1061

Maximum Entropy Charge Deconvolution in Top-down Proteomics Workflows for High Resolution Mass Spectrometry

M. L. Easterling[1], P. Sander[2], T. Roemer[2], L. DeNoyer[3], C. B. Berg[1], J. Decker[2], and M. Witt[2]

[1]Bruker Daltonics, Inc., Billerica, Massachusetts, USA; [2]Bruker Daltonik, Bremen, Germany; [3]Spectrum Square Associates, Ithaca, New York, USA

The analysis of large biomolecules from high resolution mass spectra requires reliable and fast algorithms for determination of monoisotopic masses, which are normally not detectable. For spectra acquired with electrospray ionization (ESI), the monoisotopic peak determination should be preceded by charge deconvolution to yield enhanced ion statistics by combining the peak intensities for all detected charge states.

By applying maximum entropy charge deconvolution the resultant deconvoluted spectra exhibit higher resolution, better quality, and correct isotope distributions, enabling the determination of the molecular weight with improved precision and giving an accurate information on adducts, losses or modifications.

Spectra from FT-ICR or other instruments capable of high resolving power cannot be processed with existing maximum entropy algorithms as the spectral line widths vary across the m/z range. To resolve this issue, changes to the MaxENT algorithm were made to allow processing of spectra generated by FT-ICR. The MaxENT processed spectra were then evaluated with SNAP2, a recently introduced method for monoisotopic peak assignment of high resolution data.

ESI-FTMS data from large proteins in the 30–70 kDa range were acquired. Results from the new maximum entropy charge deconvolution algorithm are compared showing improved mass resolution and isotope distribution. Protein LC-MS spectra are also shown to benefit greatly from maximum entropy processing as the peak data from multiple charge states are effectively combined for rapidly eluting peaks. Finally the result is compared with the theoretical isotope distributions for known compounds. Maximum entropy is presented within the scope of a full top-down workflow for the evaluation of intact proteins.

1062

Affinity-tagged Protein Purification in an Automated System Increases Access to Purified Recombinant Proteins

L. Usinger

Bio-Rad Laboratories, Hercules, California, USA

Affinity-tagged protein purification is an increasingly common method used for generating milligram quantities of target protein for use in functional and structural experiments. The information from these experiments is used to characterize the protein and its interactions. A new system has been developed for the automated purification and buffer exchange (desalting) of target affinity-tagged proteins. This system requires minimal user interaction to set-up, and purifies target protein in quantities of 1–100 mg. Pre-programmed methods, pre-packed cartridges, and prepared buffer kits are designed to optimize affinity purification applications. Purification results of both His- and GST-tagged proteins will be presented as examples of the automated purification. The overall working time and hands-free time will be compared with traditional techniques, such as gravity flow columns. Improvements in the speed and automation of this routine purification task translate into increased lab efficiency and allow for more time spent on functional and structural experiments. The system improves the reproducibility of this technique so that all members of a lab can produce similar quality target protein.

1064

Quantification of Cell Surface and Plasma Proteins between Normal and Malignant Breast Cells

X. Liang[1], M. Hajivandi[1], E. Ragnhildstveit[2], and M. Pope[1]

[1]Invitrogen, Carlsbad, California, USA; [2]Dynal Biotech ASA, Oslo, Norway

Cell surface and plasma membrane proteins regulate cell-cell interaction and signal transduction as well as immune response. In order to understand cellular behaviors associated with dysfunction or develop new-targeted drugs it's crucial to quantify differential expression of cell surface and plasma membrane proteins between normal and diseased states. Stable isotopic labeling with amino acids in cell culture (SILAC) is an emerging technology for this purpose that is easy to use, highly reproducible, and less susceptible to quantification bias [Ong, S.E. et al. *Mol. Cell. Proteomics* (2002) 1, 376–386]. Biotin directed affinity purifications (BDAP) have been used to characterize plasma membranes in by reducing background organelle contamination [Zhao, Y. et al. *Anal. Chem.* (2004), 76, 1817–1823], however, not in differential quantitative procedures. We report a SILAC BDAP approach to quantify differential expression of cell surface and plasma membrane proteins between normal and malignant breast cells isolated from the same patient with primary breast cancer.

Approximately 70% of proteins identified are cell surface proteins or plasma membrane proteins, including cell surface receptors, metalloproteases, trimeric G proteins, ras proteins, and ion channel proteins. The use of Streptavidin-conjugated magnetic particles (Dynabeads) resulted in 5 to 10 fold more concentrated protein sample when compared to Agarose-Streptavidin. The majority of proteins remained unchanged between normal and malignant cells. However, some matrix proteins and ion channel proteins for salts and amino acids, show increased expression levels in malignant breast cells, whereas select metalloproteases, disulfide-isomerases, and cis-trans isomerases exhibited decreased expression levels in malignant cells. Some differential expression results have been validated using Western Blotting or immunohistochemistry. Although SILAC is a powerful technique for global identification and quantification, we've found that selective isolation of plasma membrane proteins can improve specificity and coverage for tissue-specific markers that differ between normal and diseased states.

1065

Semi-quantitative Analysis of the Murine Cardiac Mitochondrial Proteome by Tandem Mass Spectrometry

X. Li, Y. Wang, C. Zong, A. Gomes, J. Zhang, and P. Ping

David Geffen School of Medicine, UCLA., Los Angeles, California, USA

Mitochondria have an established role in metabolism and cell death and have recently been recognized as signaling hubs. In the heart, mitochondrial dysfunction is linked with cell death during heart attacks through a phenomenon known as permeability transition. Although previous proteomic investigations of mitochondria have been reported, two critical pieces of information are currently lacking. First, large-scale quantitative assessment of mitochondrial proteins has never been reported from murine hearts and second, the proteins released from cardiac mitochondria during injury have never been evaluated. In this study, functionally viable murine cardiac mitochondria were purified, membranes lysed with 0.5% DDM, proteins separated by SDS-PAGE, and tryptic peptides analyzed by LC-ESI/MS/MS. Semi-quantitative information was obtained for 424 identified proteins on the basis of extracted ion chromatography (EIC) of precursor ions using the following criteria: a) only peptides unique to the given protein were used; b) the EIC peak areas of the 2 most intense peptides for a given protein were similar in magnitude; and c) peaks for semi-quantitation were chosen only within a narrow LC elution window around the MS/MS spectra yielding the highest peptide identification score. The semi-quantitation method was first validated using a standard mixture of known protein amount. Analysis of cardiac mitochondria indicated a large dynamic range of protein abundance ($\sim$4 orders of magnitude) and bell-shaped distribution of protein amount. Proteins identified from inner membrane, outer membrane and matrix were also examined to reveal relative abundance information for these sub-organelle locations. To investigate the effects of lethal or reversible injury on mitochondria, we repeated these semi-quantitative analyses on proteins released from hypotonically-ruptured or calcium-treated mitochondria, respectively. These studies reveal distinct changes in the mitochondrial proteome that accompany varying degrees of injury in this organelle and provide the basis for future investigations of mitochondria in the healthy and diseased myocardium.

1067

Proteomic Analysis of Rat Liver Microsomal Proteins by a Liquid-based Two-dimensional Chromatography Combined with Nano-LC-ESI-MS/MS

H.-J. Lee, M.-S. Kown, E.-Y. Lee, and Y.-K. Paik

Yonsei Proteome Research Center, Yonsei University, Seoul, Korea

A liquid-based two-dimensional separation system (2D-LC) method combined nano-LC-ESI-MS/MS was developed for analysis of hydrophobic rat liver microsomal proteins. The membrane proteins were separated on chromatofocusing as neutral pH range (pH 8.5–4.0) and basic range(10.5–7.4) for the first dimension followed by non porous reversed-phase HPLC for the second dimension. The detergent 2% Trion X-100 was used for efficient solubilization of membrane proteins and removal of soluble protein. Nano-LC-ESI-MS/MS was used for large-scale identification of fractionated proteins. Among 140 proteins identified by nano-LC-MS/MS, about 48% proteins were hydrophobic proteins with more than one transmembrane domain (TMD) up to proteins with 15 TMDs. Approximately 28% of identified proteins was located on membrane region. With respect to function, 28% proteins were identified to be involved in metabolism, 14% protein were related to cell signal and regulation and 14% proteins involved in transportation. The 2-D virtual image of pI vs. hydrophobicity could be used for differential display analysis. Thus, our 2D-LC system can provide large-scale analysis of hydrophobic rat liver microsomal proteins not only for the analytical separation but for the preparative separation with high resolution as compared to traditional gel methods. [Supported by MOHW grant 03-PJ10-PG6-GP01–0002 (to Y.K.P.)].

1068

Subtle Modification Isotope Ratio Proteomics; Manipulation of ^{15}N/^{14}N Ratio in Whole Plants Using Hydroponics

A. D. Laganowsky[1], J. Nishio[2], S. Bassilian[1], F. Halgand[1], P. Souda[1], K. F. Faull[1], J. Katz[3], and J. P. Whitelegge[1]

[1]University of California, Los Angeles, Los Angeles, California, USA; [2]California State University, Chico, Chico, California, USA; [3]Cedars-Sinai Medical Center, Los Angeles, California, USA

Subtle modification of isotope ratio proteomics (SMIRP) has been proposed as minimally intrusive in vivo labeling technology for human proteomics. Carbon or Nitrogen stable isotope ratios are altered minimally such that typical peptide isotopomer distributions are significantly modified, yet not so much as to dramatically extend the envelope. The isotope ratio is calculated from the intensities of individual isotopomer peaks, and knowledge of the peptide atomic formula. Where two samples have been mixed for relative expression measurements, the measured composite isotope ratio reports the relative contribution of each sample from a single isotopomer distribution.

The model plant *Arabidopsis thaliana* ecotype Columbia was grown hydroponically from seed in Hoagland's medium adjusted with 0, 1, 2 or 3% atomic ^{15}N. After five weeks of growth half the plants were subject to Fe deficiency and two weeks later intact chloroplasts isolated. Control and Fe deficient thylakoid membrane proteins from 0, 1, 2 or 3% atomic ^{15}N Hoagland's medium were digested with trypsin and subject to LC MS/MS. The subtle modification of whole plants provides slight changes in a single isotopomer distribution of peptides and ^{15}N/^{14}N ratios determined by Isosolv. The ^{15}N/^{14}N isotope ratios matched the ^{15}N atomic percentage in the hydroponic medium. Secondly, LCMS+ was performed and fractions subject to LC MS/MS, again matching the results previously described. Mixtures of 3% ^{15}N control/0% ^{15}N Fe deficient and 0% ^{15}N control/3% ^{15}N Fe deficient thylakoids were subject to LCMS+. The fractions from the mixtures, control and Fe deficient, were analyzed by LC MS/MS. The ^{15}N/^{14}N ratio of peptides was used to provide quantitation between control and Fe deficient thylakoid membrane proteins.

1069

Expression Changes in Human Salivary Proteome Using Quantitative 2D-Chromatography

P. Souda, F. Halgand, J. Loo, D. Wong, K. F. Faull, and J. P. Whitelegge

University of California Los Angeles, Los Angeles, California, USA

Human saliva represents an abundant and non-invasively available body fluid which makes it an attractive medium for biomarker discovery. Due to the complexity of salivary proteome, single dimensional separation is inadequate for separating the wide array of proteins. A two dimensional approach with mass spectrometry is critical in identifying components of the salivary proteome. Proteome Lab PF2D system was used for separating complex salivary proteomes into over 700 fractions. Parotid and whole saliva samples were obtained from the patients recruited by the clinical core of the salivary proteome project. The samples were dissolved in a denaturing lysis buffer containing non-ionic detergents, neutral chaotrope and reductants. The 2D chromatograms were compared using software Proteovue and Deltavue. The resulting fractions are analyzed by on-line LC-MS using API III+ or offline by nanospray using QSTAR-XL or LTQ-FT mass spectrometers. Hypothetical Intact Mass Tags (IMTs) generated using the QSTAR and FT-MS are used for validation using Prosight PTM. Peptide identification is done using Mascot in a bottom up fashion.

The parotid and whole saliva patterns obtained from Deltavue were uniquely distinguishable. Reproducibility of PF2D could also be characterized by fractionating identical saliva samples. Top-down collision-activated dissociation experiments have yielded useful MSMS spectra on smaller salivary proteins using both a quadrupole time-of-flight instrument and an ESI FT-MS. The 2D system can employed can also be directly interfaced to an electrospray source for LC-MS through the use of a splitter. The waste line can be directed to a fraction collector for concomitant fraction collection (LC-MS+). Fragment masses are submitted to Prosight PTM (https://prosightptm.scs.uiuc.edu) for identification of proteins and their post-translational modifications. Results are validated by parallel bottom-up experiments on the collected fractions.

1070

Expanding the Capabilities of Peptide MRM-based Assays in Plasma Using a Hybrid Triple Quadrupole Linear Ion Trap Mass Spectrometer

C. L. Hunter[1] and L. Anderson[2]

[1]Applied Biosystems, Foster City, California, USA; [2]Plasma Proteome Institute, Washington, DC, USA

As the study of protein biomarkers increases in importance, technical limitations to the detection of low abundance proteins and high-throughput, high precision quantitation remain to be overcome. The complexity and dynamic range of the plasma proteome makes the task of specific, quantitative detection even more challenging. Multiple reaction monitoring (MRM) capabilities of triple quadrupole MS systems have been explored as solutions to this challenge due to their well known sensitivity and selectivity for components in complex matrices such as plasma. Recently, a suite of over 100 MRMs representing ~50 plasma protein markers were monitored quantitatively in a single assay using the MRM-based technique showing detection of proteins down to the level of L-selectin (~1 μg/ml) with minimal sample preparation and no peptide or protein standards for most of the plasma protein markers.[1]

As more extensive candidate biomarker panels are being identified, MRM assays will need to be more rapidly developed to verify the expression changes of these proteins across larger clinical sample sets. To do this, the unique combination of triple quadrupole and ion trapping capabilities of the hybrid triple quadrupole-linear ion trap mass spectrometer have been utilized. A strategy for rapid MRM assay development for larger scale profiling and qualification of biomarker candidates without having to first prepare synthetic peptide standards is currently being investigated. A chemical labeling strategy has been employed to create global reference standards to enable quantitative comparisons between clinical samples. Assays consisting of several 100s of MRM transitions have been developed for this rapid qualification phase, facilitated by intelligent use of retention time windows during an LC analysis, while maintaining an optimum number of data points for improved precision of peak area and quantitative profiling. This presentation will demonstrate details of this workflow.

Reference

1 Anderson L and Hunter CL (2006) *Mol. Cell. Proteomics* 5.4, 573–588.

1071

Proteomic Mapping of Caveolae in Endothelium in Vivo; Distinct Functions and Signaling Pathways

A. Simonson, P. Oh, J. Yu, Y. Li, H. Witkiewicz, B. Midthune, S. Shore, L. Carver, and J. Schnitzer

Sidney Kimmel Cancer Center, San Diego, California, USA

In the same way that plasma membranes form cell surfaces that separate the cytoplasm of cells from the extracellular environment, in tissue the vascular endothelium forms a critical interface between the circulating blood and the inside of the tissue. The molecular architecture of the endothelial cell surface and its microdomains, caveolae and lipid rafts, is largely unknown. We performed comprehensive comparative proteomic analysis of the endothelial cell (EC) plasma membranes and their caveolae isolated from lung tissue. We performed trypsin digestion either directly on the membrane proteins for 2-dimensional nano-HPLC separation of peptides for tandem mass spectrometry or after protein separation by 1or 2-D gel electrophoresis. We utilized peptide mass spectra to estimate relative abundance of each protein identified in EC subfractions, and detected >300 proteins concentrated in caveolae. We are employing a comprehensive bioinformatic interrogation of these proteins to explore their Gene Ontology (GO) biological processes, and have found both expected and unexpected processes represented by proteins that are enriched in caveolae. Using various public databases, we are also uncovering new signaling pathways previously unknown in caveolae, as well as confirming known pathways. We are currently validating these findings with localization studies in caveolae involving electron microscopy, tissue immunohistochemistry and/or dual immunofluorescence confocal microscopy and undertaking functional studies of proteins of interest.

1073

The Paragon Algorithm: A Sequence Tag-based Search Engine That Substantially Improves Peptide Identification by Using Sequence Temperatures and Feature Probabilities

S. L. Seymour, I. V. Shilov, A. A. Patel, A. Loboda, W. H. Tang, S. P. Keating, and D. A. Schaeffer

Applied Biosystems|MDS Sciex, Foster City, California, USA

A novel database search engine for the identification of peptides from tandem mass spectra is presented. Compared to conventional searches, the algorithm enables routine searching of a massive effective search space (the set of modifications, cleavage events, and other peptide features allowed for) with only a small cost in actual search space (the number of specific peptide hypotheses actually scored). Sequence temperature values (STV) are computed using a sequence tag algorithm for each region of a database. Features such as modifications, substitutions, and cleavage events are all allowed for and modeled with probabilities, rather than requiring users to carefully select allowable features and considering all user-selected features equiprobable. Of the massive set of peptide hypotheses that sufficiently match an observed peptide's mass, the algorithm selects only a tiny fraction to score. This hypothesis selection is robust because it jointly considers a peptide's feature probabilities and the STV. The result is that hundreds of biological and artifact modifications, all possible amino acid substitutions, and all levels of conformance to the expected digestion pattern can be searched in a single search step in only 2–5 times the search time of conventional small search space searches with a specified digestion such as trypsin. Despite this large increase in effective search space, there is no concomitant loss of discrimination that typically comes with the exploration of large search space. This also yields a quantum improvement in searching without digestion specificity, where typical search times are an order of magnitude faster than Mascot while yielding approximately twice the peptide identifications at a fixed false positive threshold. Additionally, the use of feature probabilities enables a dramatic simplification in the user interface. The user specifies obvious details about the sample that are then translated into optimal algorithmic parameters.

1076

The Application of the Paragon Algorithm to the Study of Protein Modifications

S. L. Seymour, I. V. Shilov, A. A. Patel, A. Loboda, W. H. Tang, S. P. Keating, C. L. Hunter, and D. A. Schaeffer

Applied Biosystems|MDS Sciex, Foster City, California, USA

Protein modification studies frequently focus on phosphorylation for good reason. However, there is a great diversity of other biological modifications that also have structural and functional importance. Their detection is potentially valuable for biomarker discovery and other research. The RESID Database of biological protein modifications had 420 entries as of June 2006, yet only a small fraction of these are searched for in regular practice. This is not for lack of interest; it is a result of the lack of good analysis tools. Existing tools to search for many modifications suffer from one or more of the following problems: (1) poor discrimination; (2) excessive computational time; (3) the need for extra manual steps; (4) the inability to find previously unknown instances of modifications, or (5) the dependence on identifying a small "correct" set of proteins in a first-pass search that excludes most modifications (which reduces or eliminates the use of modified peptide identifications to refine the set of proteins detected).

The Paragon algorithm offers new possibilities in this area. This new algorithm for peptide identification uses a novel sequence tag-based approach in conjunction with probabilities for features like modifications and cleavage events such that very large numbers of modifications can be searched in regular practice without the conventional drawbacks just cited. The cost of this kind of search is only 2 to 5-fold the search time of conventional small search space where only a handful of modifications are considered. This can even be done in conjunction with searching without digest specificity—an application that is completely intractable by other means. The presentation will focus on the application of this algorithm to various samples to demonstrate the value of searching for hundreds of modifications as a regular practice.

1077

Comprehensive Difference Gel Electrophoresis (DIGE) Study Using Multivariate Analysis and ECL Plex for Verification

P. Beckett[1,2], V. Ruddat[1,2], and R. Asbury[1,2]

[1]GE Healthcare Biosciences, Piscataway, New Jersey, USA;
[2]GE Healthcare Biosciences, Sunnyvale, California, USA

This work shows a typical 2D DIGE experiment as applied to a human sample. It is a comprehensive proteomics study using prefractionation. The data analysis was performed using DeCyder 2D software including multivariate statistics in DeCyder EDA. Identifications were carried out with MS/MS on a Thermo LTQ Mass spec. Some of the proteins of interest were verified with 2D Westerns using ECL plex. This combined approach proves very successful in differential analysis.

1078

Human ex Vivo Whole Blood Infection Model for Assessing Proteomic Host Response to Pathogens

B. A. Chromy, T. H. Corzett, P. McQueen, J. Montgomery, and S. L. McCutchen-Maloney

Lawrence Livermore National Lab., Livermore, California, USA

To characterize human response to pathogens, we have developed a human ex vivo whole blood model. This host-pathogen interaction model allows the identification of early response biomarkers and provides a relevant model to characterize virulence and pathogenicity mechanisms. Briefly, blood is collected from healthy volunteers, exposed to pathogens for the ex vivo infection, and then plasma and white blood cells are characterized by multiple proteomic platforms including protein arrays, multiplexed electrophoresis, and mass spectrometry. To date, we have characterized exposure to multiple strains of *Yersinia pestis*, *Bacillus anthracis*, and several near neighbor pathogens. Notably, proteomic characterization of host response clearly discriminates 7 pathogen exposures and results are being compared to relevant, in vivo animal models. While additional studies are required, differentially expressed proteins identified following infection represent a putative biomarker panel for early detection of infectious diseases. The blood model can also be used to determine successful vaccination. Differences in proteome expression patterns were detected when comparing B. anthracis exposure in naïve and vaccinated individuals. Finally, the ex vivo model provides a novel approach to characterize virulence and pathogenicity of individual pathogen proteins and identify novel therapeutic targets by comparing host response of mutant strains with fully virulent pathogens.

1079

Antibody-Mediated Biomarker Discovery and Validation (AMBIODV)

W.-W. Zhang, L. Huang, K. Curran, and X. Fang

GenWay Biotech, Inc., San Diego, California, USA

One of currently-unmet biomedical R&D demands is to annotate gene expression profiling at protein level. Using antibody to delineate the relationship between gene expression and biomarker characteristics is one of proven effective approaches. The antibodies isolated from egg yolk are called IgY (Immunoglobulin Yolk). Despite the overall similarities between chicken IgY and mammalian IgG antibodies, there are some profound differences in their chemical structures and biological features. IgY antibodies have multiple advantages over IgG antibodies, including high specificity, strong avidity, scalable productivity, low assay background, and applicability to many immunoassays. The Fc region of IgY has an additional pair of carbohydrate chains with unique sugar residues, which is important for highly-efficient covalent coupling to solid surfaces such as microbeads or arrays. The phylogenetic distance between chickens and mammals often makes the chicken immune system far more responsive to mammalian protein antigens thus generating highly reactive antibodies. We have developed a proprietary technology that can generate libraries of gene-specific and domain-targeted polyclonal antibodies in chickens based upon bioinformatic analysis of gene-expression data and preparation of gene-specific antigens. With these antibodies, we can localize and identify unknown proteins by screening against biological samples. This approach is called "Antibody-Mediated Biomarker Discovery and Validation (AMBIODV)", which is the methodology of using antibodies to bridge the gap between gene expression and protein profiling. This process has been further developed by applying multiplex technology (antibody chips, antibody beads or tissue arrays) to identify unknown proteins in different tissue or cell samples under various physiological or pathological conditions. The IgY-mediated protein fractionation for sample preparation of proteomic studies can also greatly enhance this process. Novel proteins or biomarkers are further analyzed and validated via 2DE and mass spectrometry. Large quantities of data are generated and organized into an integrated and relational database.

1080

A Standard Nomenclature for Modifications Encountered in the Analysis of Proteins by Mass Spectrometry

S. L. Seymour[1], R. C. Beavis[2], P.-A. Binz[3], D. M. Horn[4], R. J. Chalkley[5], D. Saukau[6], L. Montecchi-Palazzi[7], J. S. Garavelli[7], J. Shofstahl[8], J. Langridge[9], D. M. Creasy[10], and J. S. Cottrell[10]

[1]Applied Biosystems|MDS Sciex, Foster City, California, USA; [2]Beavis Informatics, Winnipeg, Manitoba, Canada; [3]Swiss Institute of Bioinformatics, GeneBio, Geneva, Switzerland; [4]Agilent Technologies, Santa Clara, California, USA; [5]University of California, San Francisco, California, USA; [6]Bruker Daltonik GmbH, Bremen, Germany; [7]European Bioinformatics Institute, Cambridge, United Kingdom; [8]Thermo Electron, San Jose, California, USA; [9]Waters Corporation, Manchester, United Kingdom; [10]Matrix Science, London, United Kingdom

In the identification of proteins and peptides using mass spectrometry, artifactual and biological modifications are indicated as differential moieties on the polypeptide sequence. Because there are many different software applications that conduct identification analyses on this type of data, and there has been no standard for the nomenclature of modifications, the reconciliation of results originating from different tools can be challenging. There are currently multiple sources of modifications names, including the major public reference sites UniMod, RESID, and Delta Mass, and also software-specific catalogs. In order to remove this unnecessary complication, these authors formed a small working group that collectively represents the interests of a significant number of databases, search engines, and other software tools. By the arduous alignment of these multiple resources and point by point debate of specific names and the general principles, we arrived at a proposal for standard names of approximately 300 modifications and also a set of nomenclature principles to guide the addition of new modifications in subsequent versions. This was presented at HUPO Proteomics Standards Initiative (PSI) working group meeting in San Francisco, CA in April of this year and has since been adopted as Version 1. These standard names have been integrated as synonyms into the PSI-MOD modification ontology created by the PSI sub-group of the same name. This will allow reconciliation of modifications with other domains such as the study of molecular interactions.

1083

Advanced Tools for Glycopeptide Detection and Characterization: Incorporation of Directed Data-dependent MS3 Acquisition for Localizing Glycosylation Sites

S. M. Peterman[1,2] and K. Miller[1,2]

[1]Thermo Electron Corp., Somerset, New Jersey, USA; [2]Thermo Electron Corp., San Jose, California, USA

Glycopeptide analysis by mass spectrometry presents many problems for high-throughput protocols due to the chemical properties of the oligosaccharide modification. Oligosaccharide heterogeneity eliminates most protein database searching algorithms due the inconsistent mass shift and is further complicated by the resulting MS/MS product ion spectra dominated by low-energy sugar bond dissociation. While the resulting MS/MS spectrum enables elucidation of the glycan composition, localization of the glycosylation site can remain ambiguous due to the reduced formation of b- and y-type fragments. Our approach is to utilize the characteristic fragmentation pattern by directing MS3 scan events to select the ion containing the intact peptide backbone modified by N-acetylglucosamine resulting in a database searchable MS3 product ion spectrum rich in b- and y-type fragments.

All experiments were performed on a hybrid linear ion trap/orbitrap mass spectrometer using an experimental sequence comprised of four scan events. Scan event 1 was acquired in the orbitrap using a resolving power of 60,000 (@ m/z 400) and followed by one data-dependent MS/MS event acquired in the linear ion trap. The two remaining scan events were set to acquire data-dependent MS3 spectra on the two most intense MS/MS product ions within a user-defined mass range (800–1300 Da) to increase interrogation probabilities of the "b" ion. All protein database searching was performed using Bioworks 3.2.

Evaluation of the directed MS3 approach for glycoprotein characterization was done using fetuin and α-1-glycoproteins from bovine, human and baboons. Each sample was useful due to the high degree of glycosylation, numerous glycoforms, and abundant research in publication. Initial results identified and sequenced over ten glycopeptides in an automated fashion for each glycoprotein with greater than 50% coverage for each b- and y-type ions. A comprehensive workflow will be presented enabling complete glycoprotein identification and characterization.

1084

The Effect of Data Analysis Choices on the Number of Proteins Identified in MS/MS Proteomics Experiments

B. C. Searle[1] and P. A. Wilmarth[2]

[1]Proteome Software Inc., Portland, Oregon, USA; [2]Oregon Health and Sciences University, Portland, Oregon, USA

In MS/MS proteomics experiments, the number of identified proteins in a sample clearly depends on the quality of the separation and choice of mass spectrometer. What is not as widely known is that the number is also dependent on choices made during data analysis, which include peptide score thresholds, enzyme search specificity, choice of protein database, use of protein parsimony filters, and minimum peptide count. By juggling these choices, it is possible to "identify" widely varying numbers of proteins for the same data set. To explore this problem, we analyzed three replicates of a human salivary sample in the Scaffold data visualization software using a variety of criteria described in recent publications. For the same data set, the number of "identified" proteins ranged from 150 to 1400 depending upon the criteria used. The primary criteria affecting this variation were the use of protein parsimony and peptide count. This data suggests that any assumptions of experimental sensitivity based on the number of reported proteins are problematic given the variety of analysis protocols currently in use.

One way of comparing data sets from different experiments is through the use of protein false positive rates (FPRs) generated in reverse database searches or protein probabilities. However, most analysis techniques used in the literature focus on generating accurate peptide FPRs and ignore the fact that low peptide FPRs do not necessarily indicate low protein FPRs. Since the number of peptides identified in typical experiments can exceed the number of proteins by about a factor of 10, we show that a 1% or 2% peptide FPR can easily represent a 10% or 20% protein FPR. Finally, we demonstrate that Scaffold is a powerful analysis tool to evaluate protein identification criteria because it accurately estimates protein probabilities and protein FPRs from the most widely-used search programs.

1086

Improving Computer Interpretation of Linear Ion Trap Proteomics Data Using Scaffold

B. C. Searle and M. Turner

Proteome Software Inc., Portland, Oregon, USA

The new generation of linear ion traps (for example Thermo's LTQ) have a much shorter duty cycle than older 3-D ion traps, resulting in a five-fold increase in MS/MS scans. One side effect of faster scan rates is that while a deeper analysis is possible, substantially more uninterpretable spectra are acquired and false positives become a serious concern. To emphasize this point, in control experiments we regularly see linear ion trap data sets with less than 5% of the MS/MS spectra correctly identified, as compared to between 20% and 30% with 3-D ion traps.

The Peptide Prophet algorithm, used by the MS/MS data visualization software Scaffold, is an excellent tool to handle false positives because it assigns probabilities to peptide identifications by comparing their scores to those of clearly false matches. One major hurdle with linear ion trap data is that the increase in uninterpretable spectra places an over emphasis on false matches, which decreases the accuracy of assigned probabilities.

We have extended this method in Scaffold to handle this problem in a novel manner. First we derive a MS/MS spectrum quality filter specifically tailored to the sample. Using this quality filter, we can remove over half of the uninterpretable spectra with no reduction in sensitivity. Finally, we employ the Peptide Prophet algorithm to calculate peptide probabilities using this reduced data set. The accuracy of probabilities is increased because a large portion of uninterpretable spectra has been removed.

This reduction in data set size also results in faster computation and the ability to handle significantly more data.

1088

Protein Analysis Done at a Warp Speed; Combining Small Particle Packing Material with Nano-LC

R. J. Barkovich[1], M. Blackburn[1], E. Stover[1], S. M. Peterman[1], B. Coulombe[2], and D. Faubert[2]

[1]Thermo Electron Corp., San Jose, California, USA; [2]IRCM Proteomics Discovery Center, Montreal, Quebec, Canada

The use of small particles (<2.0 um) as a stationary phase for liquid chromatography has shown many advantages to traditional packing materials. The increase in theoretical plates at higher flow rates facilitates ballistic gradient separations with greater resolution, efficiency, and sensitivity. However, the incorporation of small particle packed LC columns into experimental protocols has been used almost exclusively for small molecule metabolomics and biotransformation studies. Minimal efforts have been devoted to the application of small particle columns to large molecule proteomics research, though protein analysis by LC/MS would clearly benefit from faster, higher efficiency separations. The inherent sample complexity associated with biomarker discovery and shotgun proteomics analyses combined with narrow peak widths afforded by high resolution small particle LC columns place extreme demands on MS experiment cycle time. For these reasons, a hybrid linear ion trap/orbitrap mass spectrometer is highly desirable.

The present research was conducted using a commercial HPLC system coupled to a hybrid linear ion trap/orbitrap mass spectrometer. Comparative chromatographic performance was accomplished using two 75 μm picofrit columns, 10 cm bed length packed with either 5 μm or 1.9 μm particle size C18 media. Protein digest samples consisted of a standard five protein digest mixture covering concentration ranges of 100 amol to 1 pmol as well as spiked human plasma digest. Mass spectral acquisition on the LTQ Orbitrap was done using parallel data acquisition enabling both peptide mass fingerprinting as well as Bioworks database searching.

Initial results will focus on comparative chromatographic performance for the two packing material based on chromatographic resolution, peptide detection sensitivity, and elution reproducibility for the simple protein digest mixture as well as the spiked human plasma digest. In addition, protein identification and sequence coverage will be presented using both Bioworks search results and peptide mass fingerprinting.

1089

A Rapid MALDI/TOF Procedure for the Analysis of Intact Hydrophobic Proteins

M. Hajivandi, X. Liang, J. Leite, and M. Pope

Invitrogen, Carlsbad, California, USA

As a class, hydrophobic proteins are translocated to cellular membranes and are critically important modulators of cellular signaling, vesicle trafficking, lipid metabolism, electron transport, pathogenic infection immune response. More than 50% of all major drug targets are membrane-bound GPCRs and ion channels and in diploid organisms, these multiunit receptors have numerous allelic forms. With the additional heterogeneity of alternate splicing and posttranslational modifications, it would clearly be difficult to disentangle the native covalent state of integral membrane proteins without peptide mapping and analysis of the intact protein's modification status. Unfortunately, MS-incompatible detergents such as SDS, CHAPS and NP-40 are usually required to maintain hydrophobic proteins in solution. Here, we present a simple and rapid alternative preparation that removes incompatible buffer components prior to intact analysis by MALDI-TOF.

Using Bacteriorhodopsin, Bovine Rhodopsin, Cytochrome P450 and nicotinic Acetylcholine Receptor complex from Torpedo californica as representative membrane proteins, we devised clean-up protocols for commercial C8 and C18 resins to create a rapid and robust procedure for removing incompatible buffer components prior to MALDI. We also relate simple procedure for trypsin digestion of the purified proteins directly on the MALDI target plates immediately after spotting. Moreover, we've shown that the sensitivity is analogous to Coomassie staining and the protocol is compatible with automation or HPLC separation of more complex membrane proteomes. Using this technique we were able to accurately characterize heterogeneity within hydrophobic protein samples, integral membrane proteins and multi-subunit receptors.

1093

Top-down, Bottom-up, and Side-to-side Proteomics

R. R. O. Loo[1], Y. Yang[1], J. Dunsmore[1], K. Williams[3], P. Boontheung[1], M. Sondej[1], H. Mouttakki[2], D. Wong[1], M. McInerney[2], R. Gunsalus[1], and J. A. Loo[1]

[1]University of California-Los Angeles, Los Angeles, California, USA; [2]University of Oklahoma, Norman, Oklahoma, USA; [3]Applied Biosystems, Foster City, California, USA

Bottom-up proteomics strategies identify proteins in complex mixtures, but discard most information revealing protein isoforms, modifications, and unanticipated processing. Top-down proteomics retains this information, providing both accurate intact mass and sequence data, but relies on compatible protein separation technologies lacking 2D gels' resolution, broad applicability, and interlaboratory reproducibility.

We have developed methods to mass analyze intact proteins embedded in dried isoelectric focusing (IEF) gels. Replacing the SDS-PAGE dimension of classical 2D-analysis by MALDI-MS (Virtual 2D Gel Electrophoresis) links intact mass measurements to discrete 2D gel spots. Once linked to specific spots, mass measurements have enduring utility, applying to every previous and subsequent 2D gel. Isoform-specific measurements accessed by 2D-PAGE (e.g., antibody blots, carbohydrate composition, synthesis/degradation rates, abundance) provide biological context for these mass measurements. The correspondence between classical and virtual 2D gel electrophoresis facilitates not only top-down and bottom-up proteomics strategies, but also a hybrid strategy, which we refer to as "side-to-side proteomics".

Proteins from less than 3 to 100 kDa were measured directly from dried isoelectric focusing gels prepared from macrophage cell cultures, human saliva, high density lipoprotein particles (HDL), and mouse embryonic tissue samples. Glycoproteins and phosphoproteins are also amenable to this approach. One strategy to gain protein identification and sequence information from the virtual gel method includes coupling the IPG gels to tandem mass spectrometry. MALDI-MS/MS was performed directly from *Syntrophus aciditrophicus* IEF gels that had been digested in-gel with trypsin. These MALDI-TOF/TOF "bottom-up" measurements provided good sequence coverage, sufficient to identify gram-negative bacterial proteins. In-source dissociation (ISD) provided bonus "top-down" capability. Even when fragmentation was limited, as observed for S. aciditrophicus protein RSYN0062, the short sequence tag obtained enabled confirmation of identity.

1094

Improved Sample Loading and MS Acquisition Methods for Identification of Protein Phosphorylation Sites

X. Li, C. Zong, A. Gomes, and P. Ping

David Geffen School of Medicine, UCLA., Los Angeles, California, USA

Protein phosphorylation is one of the most important protein modifications which is pivotal in cellular processes such as cell proliferation, differentiation and signaling. Mass spectrometry has been widely used for the protein phosphorylation analysis, however challenges still remain to perform unbiased identification of phosphorylated sites. Two inherent reasons are: a) mass spectrometry ionization and detection sensitivity is much lower for phosphopeptides versus their unmodified counter parts and b) loss of phosphate during fragmentation makes unambiguous identification of phosphorylation sites in MS/MS difficult. In this study, we used belta-casein tryptic digestion as a standard phosphoprotein and optimized the MS detection sensitivity by adding 1% phosphoric acid to the sample solution prior to sample loading; furthermore, we increased the microscan number from 1 to 3 and ion injection time from 50ms to 100ms for MS/MS spectral acquisition. A linear ion trap tandem mass spectrometer (LTQ) was operated in data-dependent neutral loss scan mode to screen for potential phosphopeptides. Our findings indicated that these methods of sample preparation and ionization improved the sensitivity of detection of belta casein phosphopeptides at least 4 fold. Using this novel strategy, we successfully identified the phosphorylated site on a recombinant proteasome subunit $\alpha2$ which had been phosphorylated in vitro by PKA and separated by SDS-PAGE. We attribute the improved sensitivity to: 1) addition of phosphoric acid to the sample prior to column loading blocks the silanol groups, thereby reducing their interactions with phosphopeptides and improving LC resolution of the phosphopeptides (as has been previously reported); 2) increasing the microscan number and ion injection time improves the spectral quality by increasing the signal to noise ratio. In conclusion, this improved method affords improved sensitivity for phosphopeptide detection on purified target proteins.

1096

Top Down Proteomics of an Integral Membrane Protein Complex, Cyanobacterial NADH Dehydrogenase

S. Bassilian[1], B. Cramer[2], H. Zhang[2], and J. Whitelegge[1]

[1]UCLA, Los Angeles, California, USA; [2]Purdue, West Lafayette, Indiana, USA

Integral membrane proteins make up around one third of the proteome yet present challenges for proteomics because of awkward physical properties including their hydrophobicity. Techniques that provide access to the transmembrane domains are particularly sought after. One way to approach this problem is with top-down proteomics whereby proteins are kept intact prior to analysis by mass spectrometry and gas-phase dissociation techniques. In this study we analyzed a functional sub-proteome, an integral membrane protein complex responsible for oxidation of NADH with electron transport to plastoquinone in the membrane.

Intact proteins in the NADH Dehydrogenase (ndh) complex of *Thermosynechoccus elongatus* were separated by reverse phase liquid chromatography with online electrospray-ionization mass spectrometry plus fraction collection (LCMS+). The intact membrane proteins produced clear mass spectra allowing mass measurement for all subunits, ranging from 4–45 kD. Intact mass tags were reported for eleven major proteins. Corresponding fractions were trypsinized and identified by micro-liquid chromatography and tandem mass spectrometry (LC-MSMS), confirming the identity of the complex and reporting hypothetical functional association of new subunits. Intact mass tags were compared to values calculated based upon translations of genomic sequence revealing post-translational modifications. Some of the subunits were further analyzed by top-down mass spectrometry, confirming the usefulness of this technique for analysis of transmembrane domains.

1097

144-Sample On-target Parallel Homogenization/Transfer Method for Solvent-free MALDI-MS Analysis of Peptides and Proteins

S. Trimpin[1], C. N. McEwen[2], H. Ji[3], and M. L. Deinzer[3]

[1]Oregon Health & Science University, Portland, Oregon, USA; [2]DuPont Corporate Center for Analytical Sciences, Wilmington, Delaware, USA; [3]Oregon State University, Corvallis, Oregon, USA

Improvements in the efficiency of peptide and protein mass analyses are constantly being made. Among the goals are better sample preparation methods, automation, and reliability. Solvent-free MALDI-methods follow this agenda demonstrating improvements in the general analyses of compounds where solvent-based MALDI-methodology encounters difficulties e.g. insolubility, analyte/matrix/solvent-incompatibility, and solvent-sensitivity relative to oxidations (Met) and other possible unforeseen reactions. Adaptations to biologically-relevant samples have made fmol amounts accessible ($200\mu L$ vials for homogenization). Current solvent-free MALDI-methods may allow making automation possible. The major limitation in the speed of performing solvent-free MALDI-analysis is the sample(s) powder transfer/adhesion to the MALDI-plate. On-target/solvent-free MALDI-preparation eliminates the transfer-step resulting in simplicity and time-savings, which are essential for proteomics (or H/D-exchange) studies. In this work, an efficient solvent-free multisample on-target-homogenization/transfer method allowed preparation of 144 samples simultaneously. A vortex device, which is commonly available in any laboratory, allows immediate access to this novel multisample on-target homogenization/transfer-method. The procedure (custom-made sample holder, $<10\ \mu L \times 144$) miniaturizes the original solvent-free methods, decreases analyte-consumption, and appears to be more reproducible because it incorporates a more standardized homogenization/transfer-step than previous solvent-free methods. The custom-made device attains a precise transfer to a specific MALDI-spot and is a tightly closed system preventing sample-to-sample-carryover. This multisample-approach automates to a great extend the sample preparation and underscores its potential for high-throughput of peptides and proteins. Simplifications such as use of commercially available disposable 386-sample holders and removal of salts, an obstacle in MALDI analysis in general, were explored. The applicability of LC-MALDI to digested proteins (e.g. Cyt.c, Thioredoxin, BSA, β-amyloid(1–42), bacteriorhodopsin) and of H/D-exchanged-in experiments (polyalanine) has been shown. This new approach may be useful not only for high throughput in membrane proteomics in general, but also for efficient H/D-measurements of proteins, a subject currently of great interest for screening protein ligands, agonists and antagonists.

1098

Quantification of Secreted Proteins between Normal and Malignant Breast Cells

X. Liang, M. Hajivandi, and M. Pope

Invitrogen, Carlsbad, California, USA

Secreted proteins, including growth factors, matrix proteins, metalloproteases, and protease inhibitors, regulate cell proliferation, migration, differentiation, and tumor invasion. Biomarkers of metastatic potential have been identified previously using comparative proteomics methods in microsomal fractions of prostrate cancer cell lines [Everly, P.A. et al *Mol. Cell. Proteomics* 2004]. In this report, we have used a SILAC approach to rapidly identify potential markers of breast cancer in a primary cell model and have validated several using Western blotting.

Cells utilize nutrients provided in culture media to synthesize cellular proteins. When cultured in media containing heavy isotopic forms of Lysine and Arginine, secreted proteins are found to completely incorporate these unique, MS-compatible labels [Oda, Y. et al. *Proc. Natl. Acad. Sci. U.S.A.* (1999) 96, 6591–6596., Ong, S.E. et al. *Mol. Cell, Proteomics* (2002) 1, 376–386, Ong, S.E., Kratchmarova, I., and Mann, M., *J. Prot. Res.* (2003), 2, 173–181]. A short incubation in serum-free media was employed to identify and quantify secreted proteins differing between malignant and normal breast cells. The majority are matrix proteins, cytoskeleton proteins, and proteins involved in metabolism and signal transduction. More than 10% of secreted proteins are proteases, metalloproteases or protease inhibitors and approximately 4.5% are growth factors, cytokines, and chemokines. Although most proteins remained unchanged between normal and malignant breast cells, a number of matrix proteins and protease inhibitors, showed decreased expression levels in malignant cells, suggesting that down regulation of protease inhibitors is essential for tumor cell migration and invasion. On the other hand, matrix proteins, such as osteoblast-specific factor 2, exhibited increased expression in malignant cells. We've also used Western Blotting or ELISA to validate differential expression of some targeted proteins. These results indicate that SILAC is a powerful technique for initial identification and quantification of secreted proteins that correlate with tumor invasiveness.

1099

MRM Analysis of iTRAQ Peptides for Validation of Changes in BG01 Stem Cells During Noggin- and BMP4-induced Differentiation

C. L. Hunter[1], C. Bramwell[1], J. Strahler[2], T. E. Gratsch[2], N. S. Leff[2], H. S. O'Shea[2], and P. C. Andrews[2]

[1]Applied Biosystems, Foster City, California, USA; [2]University of Michigan, Ann Arbor, Michigan, USA

Quantitative analysis of the response of BG01 human neural embryonic stem cells to noggin and BMP4 exposure using iTRAQ reagents on an AB/Maryland, USAS Sciex MALDI TOFTOF mass spectrometer led to identification of a number of responsive proteins. Noggin induces differentiation in this stem cell line and the goal is to understand the complete range of proteins responsive to noggin-induced differentiation. Of the many proteins identified, several demonstrated significant changes in levels, including β-III tubulin, alkaline phosphatase, reticulon-4 (NOGO), the DPYSL3 protein and DDX genes that regulate early differentiation, protein SET (involved in neurulation in amphibian), ELAV-like protein 1 (stabilizes Musashi1 mRNA in neural stem cells), and novel neuronal myosins. In addition, a number of potentially interesting proteins that underwent significant quantitative changes were identified by only a single peptide.

To verify these protein identifications, proteotypic peptides were predicted for these proteins and Multiple Reaction Monitoring (MRM) transitions were designed in silico to detect them. Using the MIDAS workflow on the AB/Maryland, USAS Sciex 4000 Q TRAP mass spectrometer, MRM-initiated MS/MS was performed on the iTRAQ reagent modified cell extracts. The MS/MS spectra generated by the MIDAS workflow analysis led to successful validation of a number of protein identifications and generation of high quality MRM transitions for the targeted peptides. These included myristoylated alanine-rich C-kinase substrate (IPI00219301), tropomyosins 2 (IPI00513698) and 4, calreticulin (IPI00020599), annexin A1 (IPI00549413) and annexin A5 (IPI00329801). These results demonstrate the facile transition from discovery mode on the MALDI TOFTOF to validation on the electrospray 4000 Q TRAP system. The expression changes of these specific proteins can now be specifically and rapidly monitored across many stem cell samples. Supported by NIH grants 1P41 RR 018627, P20 GM-069985, NS-39438, and MTTC GR-687.

1100

Systematic Comparison of Two Human Body Fluid Proteomes, Saliva and Plasma

W. Yan[1], W. Yu[1], M. Mueller[3], S. Cole[1], G. S. Omenn[2], J. Loo[1], R. Apweiler[3], and D. T. Wong[1]

[1]University of California, Los Angeles, California, USA; [2]University of Michigan, Ann Arbor, Michigan, USA; [3]European Bioinformatics Institute, Cambridge, United Kingdom

Many studies have reported characterization of the human salivary proteome (HSP) in the past few years. Recently, a core dataset for the human plasma proteome (HPP) was published. Due to the physiological significance of these two human fluids, it would be interesting to explore their proteomes and hereafter extend our knowledge. In this study, we compiled a list of HSP proteins by combining the published HSP identifications with the newly identified proteins, compared HSP with HPP through systematic analysis of their gene annotations, and built a web-based, searchable interface for the identifications. A total of 1478 HSP proteins with unique accession numbers has been identified so far. Blast was performed to cluster the HSP identifications having 100% identity over intact query sequence length. 1134 unique clusters are formed. The HSP identifications were then compared with the HPP dataset containing 15519 identifications. Based on the blast search, 432 of the HSP clusters are also presented in the HPP. Gene Ontology (GO) annotations of the HSP were compared with the published HPP core dataset. Significance of the GO distribution of HSP to HPP was evaluated by the p-value of the chi-square test. The GO component distribution for extracellular is over-represented but under-represented for membrane proteins in HSP compared to HPP ($p < 0.01$). The GO process distributions for response to stimulus and organismal physiological processes are enhanced in HSP. The GO function distributions are over-represented for antioxidant activity and structural molecular activity but under-represented in nucleic acid binding and nucleotide binding in HSP compared to HPP ($p < 0.01$). Besides, the protein identifications with the transmembrane domain are under-represented in HSP compared to HPP. The HSP identifications and its related information can be accessed through http://www.hspp.ucla.edu/.

Supported by UO1 DE16275 (UCLA) and MTTC 687 and 1U54-DA-021519 (UM).

1101

Protein Expression Profiling of Cultured Vascular Smooth Muscle Cells

C. A. Miller and D. X.-F. Deng

Agilent Technologies, Santa Clara, California, USA

Atherosclerosis is a chronic inflammatory disease characterized by lipid deposition and accumulation on vascular walls over many years. While systemic risk factors contribute to the development of this condition, atherosclerosis preferentially affects arteries and rarely veins. Differences in gene expression response to atherogenic stimuli have been found between smooth muscle cells (SMCs) from coronary artery and saphenous vein cultures. Determining the protein expression profile of these same samples is of great interest in furthering an understanding of the molecular differences in SMC phenotypes. Using a protein profiling approach that combines retention time with the high mass accuracy of TOF-based mass spectrometric analysis, a pilot study on cardiac artery SMCs demonstrated that responses were statistically different between the treatment groups. In this work, protein profile analysis has been done for both coronary artery and saphenous vein SMCs treated with stimuli including oxidized low-density lipoprotein and platelet-derived growth factor. Further, the targeted MS/MS of differentially expressed features has been done to identify these proteins. Informatics approaches used to facilitate the differential analysis for profiling applications are also described.

1102

Brain Proteomic Analysis in Murine AIDS

K. Takahashi, K. Saito, J. Masuda, M. Takemura, H. Ito, S. Fujigaki, and M. Seishima

Gifu University Graduate School of Medicine, Gifu City, Japan

Background: LP-BM5 virus-infected mice, murine AIDS, are using as a model for the HIV-related neurological dysfunction. LP-BM5-infected mice show the mnemonic abnormalities (i.e. spontaneous alternation behavior in the Y-maze and performance in the Morris water maze) and biochemical changes (i.e. cytokines, PAF, quinolinate, glutamate and AMPA-R) that produce neurologic symptomatic aspects for HIV-related neurologic dysfunction.

Purpose: To identify proteins associated with dysmnesia in this model, the expressing proteins in the infected mice brain was compared with the non-infected mice brain by two-dimensional polyacrylamide gel electrophoresis analysis (2-DE).

Method: Male mice of C57BL/6J stain were infected by intraperitoneal injection of LP-BM5. After 8w virus infection, hippocampus from LP-BM5-infected and control mice were used in this study. 2-DE was carried out in a horizontal electrophoresis system, IPG phorTM (GE healthcare bioscience) for the first-dimensional isoelectric focusing using 13 cm immobilised nonlinear pH 3–10 gradient strips. The gels were analysed by Image master 2D (GE healthcare bioscience).

Result: The expressing proteins in the LP-BM5 infected mice brain were compared with the control by Image master 2D. The number of spot detected by Image master 2D was approximately eight hundred protein spots. Seventy-five protein spots were distinguishable between control and infected mice. Most of the distinguished spots were downregulated in infected mice. These differences may be derived from the biochemical and/or behavioral changes following LP-BM5 infection in mice hippocampus.

Conclusion: It is well known that dysmnesia in this model is involved in biologically multiple active substances by virus infection. The distinguishable spot analysis could provide an important clue to define the mechanism of dysmnesia in this model.

1103

In Vivo Proteomic Mapping of Vascular Endothelium and Its Caveolae in Organs and Solid Tumors for Tissue-Specific Imaging and Therapy

P. Oh, J. Testa, Y. Li, J. Yu, P. Borgstrom, L. Carver, and J. E. Schnitzer

Sidney Kimmel Cancer Center, San Diego, California, USA

New targets and targeting strategies are needed to fulfill the promise of molecular medicine, gene therapy and even nanomedicine. Target discovery is complicated by extensive molecular diversity in normal and diseased tissues, arising from local variations in gene expression and tissue microenvironment. Moreover, most targets are poorly accessed intravenously by therapeutics because they reside deep inside tissues, beyond the barrier of the vascular endothelium. Yet the luminal surface of the vascular endothelium itself directly contacts circulating blood, providing an accessible interface for targeting in vivo. To reduce data complexity and focus on inherently accessible targets, we perfuse major organs and solid tumors, and isolate endothelial plasma membranes and their caveolae away from other tissues using a novel tissue subfractionation technique. We integrate this tissue subfractionation with multi-modal mass spectrometry, subtractive proteomic mapping, bioinformatic interrogation, and molecular imaging in vivo to identify and validate tissue- and tumor-induced endothelial targets accessible to intravenous antibodies. Monoclonal antibodies and state-of-the-art dynamic live imaging reveal rapid tissue-specific vascular targeting in vivo. Transendothelial transport of caveolae-targeting antibodies occurs within seconds (normal lung) and minutes (solid tumors) after intravenous injection. Caveolae function as active transendothelial pumps, transporting and concentrating specific molecular cargo even against a concentration gradient in the tissue interstitium. This facilitates specific delivery and tissue penetration far beyond that seen for noncaveolar antibodies, and appears to improve the effectiveness of radioimmunotherapy in a wide variety of cancers. The speed of immunotargeting and caveolar transcytosis in vivo underscores the physiological function for caveolae in transvascular exchange and capillary permeability and demonstrates the utility of targeting caveolae to enhance molecular and functional imaging as well as drug, nanoparticle, and gene delivery in vivo. This discovery and validation strategy uncovers targets that may be useful for non-invasively detecting, characterizing, treating, and monitoring many tumor types in the clinic.

1104

Study of the Native Microtubule Proteome by MALDI-TOF MS

N. Bec[1], B. Lacroix[1], P. Jouin[2], and C. Larroque[1]

[1]INSERM-CRLC Val d'Aurelle, Montpellier, France; [2]IGF, Montpellier, France

Microtubules are molecular assembly essential for organelles transport, mitotic spindle assembly and chromosome movement. Their dynamics characterized by successive growing and shortening phases which is described as "dynamic instability", is crucial for the achievement of the cell cycle. Tubulin isoforms as well as several microtubule associated proteins are involved in the control of this phenomenon. Many agents are able to perturb the dynamic equilibrium of microtubules and conduct either to the genomic instability via a missegregation of the chromosomes or to the cellular death. This last consequence is the base for the action of anti-cancer drugs that target microtubules. Tubulin itself is the most studied molecular target of some of these agents (taxanes, vincalcaloids), but conclusions on their mechanism of action remain speculative.

We have studied the proteome of native microtubules extracted from cells using temperature and lysis conditions that prevent any depolymerization process. The qualification as native microtubules was asserted by their intact capability to dissociate when exposed to low temperature (4 C). We present a quantitative analysis of the microtubule proteins prepared from untreated and curcumin treated HMEC cells using the difference in mass analysis of labeled lysines (DIMAL-K). After labelling, the protein mixture is submitted to a 1D electrophoresis followed by a MALDI-TOF MS analysis.

1105

A Physical Plasma Sub-proteome Revealed by Methods Developed for Membrane Proteins

F. Halgand, L. Yam, S. Bassilian, P. Ghasri, P. Souda, K. Faull, V. Shumaker, D. Puppione, and J. Whitelegge

UCLA, Los Angeles, California, USA

The power of proteomics lies in its ability to decipher numbers of proteomes and identify key proteins and biological pathways in the course of a better understanding of diseases. However one of the pitfalls of standard proteomics approaches is that they discard specific physical subset proteomes that requires particular methodologies for their study. The fatty acylation proteome is one of these. Since covalent protein lipidation events have been shown to play crucial roles in many cellular processes including human diseases, indirect and tedious methodologies have been developed to focus on the in vivo lipidation processes. Unfortunately, they hardly gave information about location and heterogeneity of the fatty acid moieties. Consequently, it falls to practical proteomics to develop specific and robust mass spectrometry strategies to target the fatty acylated sub-proteome.

Using LCMS+ method originally developed for membrane proteins, sub-populations of apo A-1 that were much more highly retained, and measured 264–266 Da larger than the dominant species, were detected from HDL preparations from a wide range of mammalian species. CNBr treatment allowed assignment of fatty acylation to the C-terminus of the protein (174–241, 7245 Da). Direct MSMS analyses allowed good sequence coverage for both the non acylated and acylated peptides. Both manual and software (Prosight PTM) analysis of the MSMS spectra localized stearate (266 Da) on Lys 206 or 204, for the second Apo A-1 acylated form. Some evidence showed that the first Apo A-1 acylated form was found to be heterogeneous with the presence of both stearate (264 Da) moiety on Lys 206 and with a stearate on Lys 204. Finally, for the last acylated form MSMS data only allowed narrowing the region carrying the fatty acid to the same highly conserved sequence portion spanning from AA 195 to 221.

1106

Proteomic Analysis of Soluble Proteins Enriched from Skeletal Muscle

M. H. Simonian, K. S. Rau, C. M. Markl, and C. K. Ratnayake

Beckman Coulter, Inc., Fullerton, California, USA

The presence of highly-abundant proteins in certain biological samples can mask the detection of lower-abundant proteins, which are potential biomarkers. In skeletal muscle tissue, the highly-abundant proteins are primarily contained in the myofilaments and consist of the contractile proteins actin, myosin, tropomyosin and troponin. This paper presents the proteomic analysis of the soluble proteins enriched from bovine skeletal muscle tissue that utilizes two-dimensional liquid fractionation. First, extraction of the soluble proteins with minimal presence of the highly-abundant contractile proteins was accomplished by homogenizing the tissue in a hypotonic buffered solution. This was followed by denaturation of the soluble protein extract. The intact denatured proteins were then fractionated by pH followed by hydrophobicity. This fractionation generates a two-dimensional, pH/hydrophobicity proteome profile. From 500 mg of tissue, the yield of soluble protein was 4–5 mg. With this two-dimensional liquid fractionation technique, over 1700 protein peaks were obtained from the soluble skeletal muscle extracts in each of three sample preparations. Select second-dimension fractions with differing amount of proteins were digested with trypsin and the peptide products were analyzed by mass spectrometry to identify the proteins. Using a simple enrichment technique coupled with two-dimensional fractionation allows lesser abundant proteins from skeletal muscle to be investigated.

1108

Consensus Search Approaches for Protein Identification by Tandem Mass Spectrometry

S. Yin[1], R. Menon[1], T. W. Blackwell[1], K. Hung[2], D. Sarracino[2], and D. J. States[1]

[1]University of Michigan, Ann Arbor, Michigan, USA; [2]Harvard-Partners, Boston, Massachusetts, USA

We used Open Mass Spectrometry Search Algorithm (OMSSA) and X!Tandem in combination to evaluate the performance of consensus peptide identification in complex mixtures. Mass spectra were obtained by LC-MS/MS analysis of human and mouse plasma and interrogated by database search to a non-redundant protein sequence collection. All measures of match quality and significance were higher for spectra that matched the same peptide sequence using both programs even when compared to single program matches of peptides derived from plasma proteins independently identified with high confidence [States et al, *Nat. Biotechol.* 24(3):333–8]. The fraction of false positives (matches to reverse sequences) was reduced from 7.1% (all OMSSA matches) to 3.8% for consensus identifications. Consensus peptide identifications matched an average of 39 and 37 different spectra by OMSSA and Tandem, respectively. Peptides identified only by OMSSA or X!Tandem matched an average of 28 spectra and 7 spectra, respectively. The total ion current of spectra matching peptides showed a similar trend. The large number of matching spectra and higher peak parent ion current suggest that consensus identifications are more abundant proteins. 76% of the peptides matched by OMSSA were also identified by X!Tandem. In contrast, only 8% of the peptides that X!Tandem identified were also matched by OMSSA. 28% of X!Tandem matches included post-translational modifications, but none of the consensus identifications contained post-translational modifications. Measures of spectrum quality such as average peak intensity, spectral entropy or peak spacing based measures did not correlate strongly with the ability of X!Tandem and OMSSA to achieve a consensus identification of a peptide sequence. In summary, the use of consensus identification by OMSSA and X!Tandem improved specificity and selects peptide identifications that have very high quality matches by either program's criteria. [This work was supported in part by grants R01LM008106, U54DA021519, P41RR018627, MTTC687.]

1109

Phosphoproteome Profile of Human Liver Chang's Cell Based on 2-DE and Fluorescence Staining

J. Liu, Y. Cai, J. Wang, Q. Zhou, L. Jiao, W. Ying, S. Sui, Y. Jiang, and X. Qian

Beijing Proteome Research Center, Beijing, China

Since reversible protein phosphorylation plays a central role in biological regulation, many methods have been developed for the systematic parallel analysis of the phosphorylation status of large sets of proteins involved in the regulatory circuitry of cells and tissues these years. To directly survey the modification extent and distribution state of the phosphoproteins in the biosystem, in this study we applied a phosphoprotein staining method with the dye of Pro-Q Diamond for high-throughput identification of phosphoproteins. The specificity of method was validated with standard proteins and the method was subsequently applied on the phosphorylated protein analysis of human liver Chang's cells. The total proteins of Chang's cells were separated on the two-dimensional gel, then sequential stained by Pro-Q Diamond and Coomassie Blue G-250. After image analysis the gel spots of phosphoproteins were cut, in gel digested and identified by MALDI-TOF/TOF-MS analysis. A total of 269 phosphoproteins were identified, and among them 30 were known phosphoproteins and embodied in the swissprot database. Simultaneously, by comparing the relative volume of the phosphoprotein map and the total protein map, we could observe the extent of protein phosphorylation. Further more, the data also showed that the phosphoprotein staining method combined with the 2-dimensional electrophoresis can detect the polymorphism of the phosphoprotein, and the phosphoprotein staining method combined with the total protein staining method can tell the high-abundance but lower phosphorylated proteins from the low-abundance but highly phosphorylated proteins. Through the analysis, we concluded that the phosphoprotein staining methods could be used for global quantitative phosphorylation detection.

1110

Identification and Analysis of Differential Humoral Response Targets in Prostate Cancer

B. S. Taylor, A. Sreekumar, M. Pal, J. Yu, R. Shen, R. Zhao, J. T. Wei, D. Ghosh, D. M. Lubman, and A. M. Chinnaiyan

University of Michigan, Ann Arbor, Michigan, USA

Limitations in current front-line methods such as prostate-specific antigen (PSA) for the early detection of prostate cancer (PCa) from benign hyperplastic conditions (BPH) are motivating a transition to leveraging immune system-driven amplification of the autoantibody response to intracellular antigens which is more sensitive in the detection of low abundant biofluid-derived proteomic tumor markers. Here we interrogate humoral response in a serum cohort of 34 patients with either BPH or PCa with a strategy that couples multi-dimensional liquid-phase protein fractionation of localized and metastatic prostate cancer tissue lysates to protein microarrays and subsequent mass spectrometry. A supervised learning analysis of the humoral response arrays generated a 20-fraction predictor having 77.78% sensitivity and 75% specificity in predicting PCa from BPH. A bipartite signature emerges from this 20-fraction predictor, a subset of which is antigens with increase reactivity in BPH patients relative to PCa and another signature reflecting the reverse. The proteins in the signatures are identified using mass spectrometry. A molecular 'concept' analysis of the protein content of the two signatures reveals interesting patterns of enrichment for humoral response targets. This includes a validatory positive control of multiple cancer gene expression signatures representing aggressive, poorly differentiated epithelial carcinomas. Additionally, it is significant that a subset of the proteins eliciting this humoral response pattern mapped to its set of concordantly expressing genes in an independent DNA microarray study of BPH vs PCa (OR = 1.73, P-value = 0.009). Further, several of the proteins in this enrichment have known involvement in inflammatory processes and immuno-modulating behavior. The success of this method offers a means for comprehensive analysis of the cancer proteome using small quantities of analyte obtained from fractionation of proteins as expressed in cancer cells while maintaining their post-translational modifications, which are often critical to elicited humoral response profiles.

1112

MSMS MALDI Mass Spectrometric Imaging for Direct Sequence Identification of Native Neuropeptides

A. James[1], M. Willetts[2], F. Abdi[2], and D. Gostick[2]

[1]MDS Sciex, Concord, Ontario, Canada; [2]Applied Biosystems, Framingham, Massachusetts, USA

Matrix assisted laser desorption/ionization mass spectrometric imaging (MALDI-MSI) is a technique that is increasingly used to facilitate biomarker discovery and assist in drug development. Experiments based on TOFMS profiling are typically used to identify lead compounds. However, confident identification of biomolecules based solely on MS and MSMS imaging data has been a significant challenge in MSI experiments, particularly for larger biological compounds such as peptides and proteins. Here we report the sequence identification of a neuropeptide directly from a coronal rat brain section. Neuropeptide leads were identified using an image profiling experiment, followed by directed MSMS imaging experiments. The resulting MSMS imaging data provided sufficient sequence information to identify a peptide clip from the opioid precursor protein Prodynorphin. This neuropeptide was localized within the substantia nigra region of the brain and is likely a neurotransmitter. The MS profiling and MSMS imaging experiments were run on a 4800 MALDI TOF/TOF™ Analyzer

1114

Correlated Evolution of Mammalian Plasma Protein Sequences and Abundances Based on Comparative Analysis of the Human and Mouse Plasma Proteomes

R. Menon[1], T. W. Blackwell[1], K. E. Hung[2], B. Krastins[2], D. A. Sarracino[2], G. S. Omenn[1], and D. J. States[1]

[1]University of Michigan, Ann Arbor, Michigan, USA; [2]Harvard Medical School-Partners Healthcare Center, Boston, Massachusetts, USA

We identified orthologous proteins in human and mouse plasma, using tandem mass spectroscopy of tryptic digests of normal human and mouse plasma. Samples were analyzed with an LCQ instrument and X!Tandem to search the custom-made, species-specific databases based on the NCBI RefSeq database. With a Poisson model, we found 93 orthologous pairs containing 80 human proteins orthologous to 90 mouse proteins with confidence >95%. The numbers of identifying peptides are strongly correlated in the two species ($R = 0.9$). We looked for evidence, based on number of peptide matches identified, for a difference in abundance of the protein in the two species. Using Fisher's exact test, 16 orthologous pairs demonstrate a significant difference after correction for multiple hypothesis testing. The sequence divergence for plasma orthologous proteins (fractional identity = 0.72) is greater than for all human mouse (HM) orthologous pairs (fid = 0.79). The ratio of non-synonymous to synonymous mutations (0.21) is also higher in plasma orthologs than for all human/mouse orthologs (0.13). The divergence of those orthologs demonstrating a difference in abundance based on number of peptides is still greater (fid = 0.68). We developed a linear regression model based on features like total ion current and total number of scans, to predict the protein concentration. Using this model we found that albumin, serotransferrin, complement c3, hemopexin were among the highly abundant proteins in the dataset. The major protein groups in the high confident orthologous protein set were apolipoproteins, complement proteins, serine protease inhibitors and collagens. The synonymous and non synonymous rates of some apolipoproteins and serpins were high along with lower percentage identities. In summary, the low fractional identities and relatively high dNdS ratios of some blood proteins suggest an evolutionary trend in these human blood proteins.

1115

Identification of Secretion Regulatory Molecules in HBV by Glycoproteomic Techniques

T. Nakagawa[1], N. Koyama[2], E. Miyoshi[1], M. Nakano[1], N. Taniguchi[1], and A. Kondo[1]

[1]Osaka University, Suita, Japan; [2]Takara Bio. Inc., Otsu, Japan

Hepatitis B virus (HBV) is the human member of the family Hepadnaviridae and worldwide is associated with more than 350 million chronic infections and nearly one million deaths annually. HBV-associated liver diseases are now one of the important health problems in the world because of the high numbers of patients and the serious consequences.

The cell line HB611 established by transfecting the HBV genome into a human hepatoblastoma cell line Huh6 produces HBsAg, HBeAg, and HBV virion into the medium. Several studies on HBV replication or expression of HBV-related proteins have been reported using the HB611 cell. We found that overexpression of N-Acetylglucosaminyltransferase III (GnT-III) suppressed gene expression of HBV in the HB611 cells.

In this presentation, we will show that GnT-III overexpression course inhibit the secretion of HBV. And we also identified some kinds of integrins and cadherin, which had bisecting GlcNAc introduced N-glycan, played the important role of secretion of HBV by using glycoproteomic techniques.

1116

Identifying Novel Genes in the Human Genome Using Tandem Mass Spectra

D. Fermin, B. B. Allen, T. W. Blackwell, R. Menon, M. Adamski, P. Ulintz, G. S. Omenn, and D. J. States

University of Michigan, Ann Arbor, Michigan, USA

Defining genes and gene products remains a challenge. Tandem mass spectroscopy provides a novel unbiased source of data that can be interrogated using genomic sequence to identify novel protein coding sequences. A six-frame translation of the Human genome was used as the query database to search for novel blood proteins in the data from the HUPO PPP.

Significance is assessed using a Poisson statistical model incorporating the length of the matching sequence and the frequency of spectrum matches observed in searching the database [*Nat. Biotechnol.* 2006 24(3): 333–8]. Matches are binned by X!Tandem hyperscore, and statistics for each score class are considered independently. The overall probability that the matches to an ORF occurred at random is calculated as the product of the probability that the matches in each score category occurred at random. The expected number of random matches, E, is calculated as the product of the probability that an ORF match occurred at random multiplied by the number of ORFs searched. The confidence in an ORF identification is $1/(1+E)$. An open reading frame is considered significant if confidence is greater than 95%. Expanding recently published work [*Genome Biol* 2006; 7(4):R35], we have identified 837 significant open reading frames coding for 18852 peptides falling within 914 exons of 413 genes. Out of 8856 candidate ORFs outside the boundaries of known genes, 3246 of them achieved a confidence $> = 0.95$. We also required the XG ORFs to be supported by at least 3 distinct ESTs. Twenty four of the XG ORFs were found to have a significant alignment to the mouse genome. Of these, 13 of the alignments encompassed a coding region for one of the diagnostic peptides associated with the ORF. Gene models for the XG ORFS were derived from the GENSCAN prediction made for their coding regions.

1119

Biomarker Discovery Enhanced by IgY-Immunoaffinity Fractionation

X. Fang[1], L. Huang[1], N. Zou[1], P. Gagné[2], G. Poirier[2], D. Hinerfeld[3], S. Tam[3], and W.-W. Zhang[1]

[1]GenWay Biotech, Inc., San Diego, California, USA; [2]Laval University, Québec, Canada; [3]University of Massachusetts Medical School, Shrewsbury, Massachusetts, USA

Avian polyclonal IgY (Immunoglobulin Yolk) antibodies have unique and advantageous features that allow for highly-specific and effective capture of protein targets. Previously, we have reported development and application of IgY microbeads for the one-step removal of 12 highly-abundant proteins (HAP) from plasma using immunoaffinity columns. However, removal of top 12 HAP only reduces dynamic range of plasma protein concentration by 1–2 logs. The next level of abundant protein, moderately abundant proteins (MAP), becomes an obstacle to access low abundant proteins (LAP), where the majority of biologically interesting and clinically important biomarkers reside. Therefore, isolation of MAP is a new challenge for effectively and accurately detecting and analyzing LAP. To tackle this challenge, we further developed our IgY microbeads system in two aspects: (1) increasing the capacity of IgY12 column to allow more sample loading and (2) establishing SuperMix column to separate MAP from LAP. The high-capacity IgY12 column was successfully developed by modifying IgY microbeads. The SuperMix column was also successfully developed by immunizing chickens with a flow-through fraction of IgY12 column and constructing the column with affinity-purified IgY antibodies against the flow-through proteins of IgY12. The SuperMix column was utilized further to fractionate the flow-through fraction of IgY12. This resulted in an eluted fraction (designated MAP fraction) and the flow-through fraction (designated LAP fraction), which were analyzed by SDS-PAGE, 2DE and LC/MS/MS. Total 211 proteins were identified in the eluted fraction of SuperMix column (81 proteins of high-confidence identification and 130 proteins of low-confidence identification). And total 324 proteins were identified in the flow-through fraction of SuperMix column (126 proteins of high-confidence identification and 198 proteins of low-confidence identification). With high-capacity loading and coupling of IgY12 and SuperMix columns, the further development of IgY microbeads system provides an enhanced methodology to dig deeper into the plasma proteome.

1120

Top-down Identification and Characterization of Intact Human Saliva Proteins

J. P. Whitelegge[1], F. Halgand[1], S. Bassilian[1], J. Zhang[2], P. Souda[1], V. Zabrouskov[2], T. C. McClure[2], J. A. Loo[1], K. F. Faull[1], and D. T. Wong[1]

[1]University of California, Los Angeles, California, USA; [2]Thermo Electron Corporation, San Jose, California, USA

Complementary to bottom-up approaches, top-down proteomics is being applied as a powerful method to provide complete characterization of protein primary structure of proteins, including post-translational modifications. As a result, description of protein isoforms, and intra and inter individual polymorphisms would provide a better understanding of structures to functions relationship of proteins and the discovery of disease biomarkers.

Whole saliva and stratified parotid gland secretions collected from human subjects were analyzed by LC-MS+. Collected fractions were then subjected to tandem mass spectrometry using both quadrupole time-of-flight (QSTAR XL) and hybrid linear ion-trap Fourier-transform ion cyclotron resonance (LTQ-FT) mass spectrometers. Peak lists from intact protein tandem mass spectra were data mined (Prosight PTM), firstly to identify parent open reading frames, and secondly to assign primary structure, including post-translational modifications. Candidate sequences were manually processed and matched to peak list data using Single Protein Mode. In this way several saliva proteins were clearly identified by Top Down MSMS experiments. These proteins include peptide C, protein IB-8c which contains peptide F, proteins IB-6, IB-9, II-2, IB-1, PRP-3, PRP-1, Db-f, Db-s and cystatin SA1. These experiments revealed new post translational modifications in protein II-2 and cystatin SA1. Unexpectedly, isomeric sequence changes were found for peptide C and PRP-3.

Salivary proteins present a significant challenge for protein identification via top-down strategies because they are often proteolytically processed at both N- and C-termini and contain a high rate of sequence repeat unit, such that automatic matching of peak lists to unmodified translations of open reading frames are unsatisfactory. Furthermore, the diversity of potential sequence polymorphisms and post-translational modifications precluded generation of (prejudiced) shotgun-annotated databases. The conclusion of this study highlights the ultimate requirement for unprejudiced interpretation of tandem mass spectral data with the view to discover new biomarkers of human pathologies in saliva.