

Probability-Based Protein Identification for Post-Translation Modifications and Amino Acid Variants Using Peptide Mass Fingerprint Data

Weiwei Tong (1,2), Mark E. McComb (1), David Perlman (1), Hua Huang (1), Peter B. O'Connor (1),
Catherine E. Costello (1), Zhiping Weng (2)

(1) Boston University School of Medicine, (2) Boston University, Boston, MA

Novel Aspects: A free web based search engine with a novel algorithm for peptide mapping featuring PTM and amino acid variant search.

Introduction: Mass spectrometry (MS) has become a popular method in proteomics. We present an enhanced version of the probability-based peptide mass fingerprint database search algorithm, BUPID, to interpret MS data. BUPID applies a robust and accurate statistical model to identify proteins and find post-translational modifications (PTMs) and/or amino acid variants.

Methods: We apply a log-likelihood ratio calculation to determine the probability that a protein is present within the sample. The model distinguishes the null hypothesis: a set of peaks in the spectrum is generated by the random background, from the alternative hypothesis: the same set of peaks is generated by peptides corresponding to a specific protein. Peak assignment is based on the log-likelihood ratio as oppose to matching peaks with peptides within the mass tolerance. The mass spectrum is first matched with unmodified protein sequences. The top scored proteins in the search result are then used to construct a database that includes PTMs and amino acid variants, against which the spectrum is searched again. Results are ranked by their probability scores.

Preliminary Data: We tested the BUPID web server on 30 sets of peptide MS data. Test sets (consisting of MALDI MS obtained on proteolytic digests of proteins) were obtained from both ongoing projects at BUSM and artificially generated mass spectra. Every sample contained proteins with post-translational modifications and/or amino acid variants present within mixtures of other proteins. The MS data was first searched against protein sequences in the SwissProt database in their unmodified forms. In all 30 cases, proteins of interest were identified within the top 10 returned hits, ranked by the probability score of log-likelihood ratios. Around 500 sequences with high probability scores and sufficient sequence coverage were passed on to the second round of search for PTMs and amino acid variants. Due to the limitation of using MS spectra (only measures peptide masses) rather than MS/MS data, the algorithm was unable to single out the exact residue which may be modified if multiple copies of the same residue are present in one peptide. However, in 30% of the cases, BUPID correctly identified the PTM, with the highest probability score. In 91% of the cases, BUPID found the correct modification in the correct peptide within the top 10 hits. In all cases when it searched for amino acid sequence variations, BUPID was able to pinpoint the correct mutation as the top returned result among the set of potential mutations. The following parameters were used in the search - Taxonomy: Human; Enzyme: Trypsin; Maximum Missed Cleavage: 1; Peptide Mass Tolerance: 0.2 Dalton. PTMs: 50 common PTMs chosen from the Delta Mass database. One variant and/or 5 PTMs in the sequence were allowed during the search. A typical BUPID database search takes 2~3 minutes for the first-pass search and 1 minute for the second-pass search on a Pentium IV PC.

Acknowledgements: This project was funded by NHLBI contract N01 HV-28178.

Mapping the Stimulus-specific Signaling Pathways Involved in THP-1 Cells Exposed to Porphyromonas gingivalis LPS vs. Fimbria vs. Live P.g.

Julian A. Saba (1); David H. Perlmanm (2); Hua Huang (2); Mark E. McComb (2);
Catherine E. Costello (2); Salomon Amar (1)

1) BU Goldman School of Dental Medicine, 2) Boston University School of Medicine, Boston, MA

Novel Aspect: THP-1 monocytes sense live P.g. differently than its bacterial components.

Introduction: Periodontitis is an inflammatory disease with host-parasite interactions which contribute to connective tissue destruction and alveolar bone resorption. *Porphyromonas gingivalis* (P.g.) a black-pigmented Gram-negative anaerobic bacterium, is a major pathogen in the development and progression of periodontitis. Cell wall and appendage structures, such as lipopolysaccharide (LPS) and fimbriae, play important roles in the induction of innate immune responses, including cytokine production by localized and circulating monocyte/macrophage. Based on preliminary data, live P.g. stimulates unique pro-inflammatory signal transduction pathways in human peripheral blood monocytes (PBM) and THP-1 as opposed to purified bacterial components. In these experiments we test at the level of protein expression that unique signaling pathways are differentially induced by live P.g., LPS and fimbriae.

Methods: Human PBM (>95% pure) freshly elutriated from healthy donor and monocyte-like THP-1 cells were exposed to Live P.g., P.g. LPS and P.g. fimbriae. 2-dimensional gel electrophoresis (2-DGE) were performed on the cell lysate proteins of PBM and THP-1 differentially produced. The expressed proteins were excised and subjected to tryptic in-gel digestion. Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) and reversed phase high performance liquid chromatography (RP-HPLC) in combination with MS were performed. MS/MS is being performed to characterize post-translational modifications.

Preliminary results: We had previously shown using a proteomic approach that host immune cells sense live P.g. and its components differently. Our results reveal unique protein expressions in response to live P.g. as compared to its bacterial components. We are currently investigating, at the levels of protein expression, to determine how unique signaling pathways are differentially induced by live P.g.

Analysis of the 2-DGE profiles has led to the identification of several differentially expressed protein spots. Treatment with P.g. LPS and P.g. fimbriae resulted in differential expression of eleven and eight protein spots respectively relative to unstimulated controls. Spots of interest were excised, in-gel digested with trypsin, and analyzed via MALDI-TOF MS for protein identification. Our experiments demonstrate that the proteomics approach holds promise for revealing unique signaling pathways that are differentially induced by LPS and fimbriae. The data generated in these experiments using 2-DGE provided qualitative comparison of the proteins expressed in response to exposure to LPS and fimbriae, as well as the identification of several proteins. LPS stimulation of THP-1 cells up-regulated the expression of the following proteins compared to control: deoxyribonuclease, actin, carbonic anhydrase 2, alpha enolase, GRAP 2 human, PDI, grp78 and HSP70. In contrast, fimbriae treatment did not result in statistically significant changes to protein levels relative to control. The changes seen after fimbriae exposure were small and down-regulatory relative to controls. This suggests that comparing LPS to fimbriae is no different from comparing LPS to control. Thus, its not fimbriae that is involved in the critical role of up-regulating the innate immune response but LPS. Similar approaches will be undertaken for human PBM.

Acknowledgments: This research is supported by the National Institutes of Health Grants R01 NIH/NIDCR R01 DE015989 and P41 RR10888 and S10 15943 and NHLBI contract N01 HV28178.

A Proteomic Approach to the Study of Systemic Amyloidosis

Francesca Lavatelli, David H. Perlman, Mark E. McComb, Roger Theberge, David C. Seldin, Lawreen H. Connors, Giampaolo Merlini, Martha Skinner, Catherine E. Costello

Boston University School of Medicine, Boston, MA, and IRCCS Policlinico San Matteo, Pavia, Italy

Novel Aspect: Proteomic investigations into amyloid infiltration of adipose tissue. Novel diagnostic tool for molecular typing of systemic amyloidoses.

Introduction: Protein deposition as amyloid is the basis of diseases that, overall, have an enormous social and medical impact. More than twenty different proteins are known to be causative agents. In systemic forms, amyloid deposition is associated with dysfunction of vital organs, and molecular typing of the deposits is necessary for diagnosis and treatment. The traditional diagnostic approach is multidisciplinary, but sometimes fails to identify the correct type. A proteomic approach could help in diagnosis, through the direct molecular characterization of fibrillar proteins in tissues, and cast insights into the mechanisms of tissue damage. We are working to characterize amyloid deposits in abdominal subcutaneous fat from patients with systemic amyloidosis using 2D-PAGE followed by MALDI-TOF MS and peptide mass fingerprinting.

Method: Fat tissue samples were obtained, with institutional human studies approval, from individuals affected by various forms of systemic amyloidoses and from unaffected volunteers were used as normal controls. Protein was extracted from fat tissue by homogenization directly in IEF buffer followed by ultracentrifugation to clear debris and delipidate samples. Samples were then subjected to 2D-PAGE analysis and Coomassie or silver staining. Protein spots were imaged and quantitated using PDQuestTM software. Spots indicating differentially expressed proteins were excised and subjected to in-gel digestion by trypsin or other proteases. Peptides were analyzed by MALDI-TOF MS or by LC-MS and MS/MS. Spectra were analyzed with MoverZTM or MassLynxTM and ProteinLynxTM software, and peptide mass fingerprinting analysis utilized MASCOT and/or BUPID.

Preliminary data: A rapid methodology to prepare samples for high-grade 2D-PAGE analysis from fat tissue was developed. Samples from unaffected (non-amyloid) volunteers were used to generate 2D reference maps for comparison with those generated from patients affected by systemic amyloidoses. Preliminary results arising from the comparison of these maps indicate that significant differences exist. Characteristics of the gels from patients with systemic amyloidoses that differed from controls fell into three categories: (1) proteins were observed in regions of the gels from patient samples consistent with the expected migration of amyloidogenic proteins; (2) apparent up-regulation of some proteins in patients samples was observed, as compared to those same proteins in control samples; (3) new and unidentified spots were observed in patient samples that were absent in the controls, suggesting the appearance/recruitment of novel proteins. These new spots often appeared in “trains” suggesting the presence of modified forms of the same protein. Identification of the proteins responsible for these differences is being accomplished using in-gel protease digestion and a variety of MS techniques.

The use of 2D-PAGE as a tool to highlight the differences between diseased and normal states is well known. We are developing a proteomic approach using this technique in the analysis of fat tissues that should be able to provide a reliable diagnosis for patients in whom amyloid disease is suspected. This approach is practical and feasible, given that fat aspirates of potential amyloid patients are routinely acquired for histological analysis. Ongoing analyses may provide identification of new aspects of the disease mechanisms, including the involvement of novel proteins and protein post-translational modifications.

Acknowledgments: This research is supported by the US National Inst. of Health grants P41 RR10888, S10 RR15942, HL 68705, contract N01 HV28178 (CEC), the Amyloid Research Program at Boston University, Grants from the Italian Ministry of Health and Cariplo Foundation (GM) and the BUSM Amyloid Program. FL is supported in part by an investigator fellowship from Collegio Ghislieri, Pavia, Italy.

**Profiles of Cardioprotection: Integrated Proteomic and Metabonomic
Study of the Effects of Nitrite Treatment on the Heart**

David H. Perlman (1), Selena Bauer (2), Hua Huang (1), Maria Francisca Garcia-Saura (2), Nathan S. Bryan (2),
Bernadette O. Fernandez (2), Mark E. McComb (1), Martin Feelisch (2), and Catherine E. Costello (1)

1) Cardiovascular Proteomics Center, 2) Whittaker Cardiovascular Institute,
Boston University School of Medicine, Boston, MA.

Novel Aspect: Integrated proteomic/metabonomic study of the cardioprotective effects of nitrite, including the discovery of significant, nitrite-induced alterations to the cardiac proteome.

Introduction: Nitrite, stable and abundant in vivo, was thought for decades to be biologically inert at physiological concentrations. Formed endogenously through oxidation of nitric oxide (a ubiquitous signaling molecule known to elicit broad biological effects) and derived extensively from dietary sources, it has very recently proven to be a signaling molecule in its own right, affecting soluble guanylyl cyclase and cytochrome P450 activities, heat shock protein 70 and heme oxygenase-1 expression, and protecting against cardiac ischemia-reperfusion injury. Based on this recent evidence, we sought to characterize the impact of changes in systemic nitrite availability on the cardiac proteome, using 2D gel-based separation followed by LC-MS/MS or MALDI-TOF MS analyses, then correlate these data to metabonomic studies.

Method: Male Wistar rats administered a single intraperitoneal injection of sodium nitrite (0.1, 1 and 10 mg/kg) were anesthetized 24 hours later, perfused free of blood and their cardiac tissue was harvested. Immediately after homogenization the concentrations of NO-related metabolites were determined by gas phase chemiluminescence and HPLC. Tissue aliquots were denatured in IEF buffer and subjected to 2D-PAGE analysis, followed by Coomassie or silver staining. Protein spots were imaged and quantitated using PDQuestTM software, excised and subjected to in-gel trypsin digestion. Peptides were eluted, de-salted and analyzed by MALDI-TOF MS or by LC-MS and MS/MS. Spectra were analyzed with MoverZTM or MassLynxTM and ProteinLynxTM software, and peptide mass fingerprinting analysis was conducted using MASCOT.

Preliminary data: Using 2D gel electrophoresis, we have created preliminary 2D reference maps of cardiac proteomes of control Wistar rats and those systemically administered physiologic and therapeutic levels of nitrite. We have confirmed the identity of over a hundred isolated protein spots through in-gel digestion followed by MALDI-TOF MS and peptide mass fingerprinting or by LC-MS/MS analyses. Quantitative comparative analyses have revealed significant changes to cardiac protein expression upon treatment with nitrite. These changes consisted of both up- and down-regulation of steady-state protein levels, as well as apparent alterations in post-translational protein modification and have included proteins involved in cell structure, energy metabolism, redox balance, and chaperone activity, among others. Additionally, we have measured metabonomic changes and changes to the cellular redox status that we have correlated to these proteomics data. Specifically, we have detected changes in the ratio of reduced to oxidized ascorbate and glutathione, as well as changes in the levels of S-nitroso, N-nitroso, and heme-nitroso species in the tissues. The results obtained using this combined proteomics/metabonomics approach indicate specific changes in expression of cellular stress response proteins triggered by nitrite that may confer resistance to further oxidative insults. Our data suggest that nitrite, due to its stability in vivo, may be able to act in a long-range endocrine fashion to establish this protective tone in the heart. Our integrated proteomic and metabonomic approach is a step toward elucidating the scope and mechanism of cardioprotection and the potential activity of nitrite, a biological compound with newly discovered importance in human health and medicine. Our data may have direct and immediate implications for current experimental therapeutic uses of nitrite and may provoke a reassessment of the impact daily dietary intake of nitrite.

Acknowledgements: This project was funded by NIH grants P41-RR10888 and S10-RR15942 and NHLBI contract N01-HV-28178.

The Proteomics of Proliferation and Malignancy: Analysis of Tumor-specific and Proliferation-specific Transcription Complexes in Normal and Malignant B Cells

Paul B. Romesser, David Perlman, Anupama Sinha, Mark E. McComb, Douglas V. Faller,
Catherine E. Costello and Gerald V. Denis
Boston University School of Medicine, Boston, MA

Novel aspect: Proteomics of resting, proliferating, and lymphomic B cells. New identifications represent important potential targets for chemotherapeutic intervention in non-Hodgkin's lymphoma.

Introduction: The dual bromodomain protein Brd2 is closely related to the basal transcription factor TAFII250, which is essential for cyclin A transactivation and mammalian cell cycle progression. In transgenic mice, constitutive lymphoid expression of Brd2 causes diffuse large B cell lymphoma, a common form of non-Hodgkin's lymphoma, in part through constitutive transactivation of the cyclin A locus. In an effort to understand the mechanism underlying Brd2-driven B cell lymphoma and how it differs from proliferating or resting B cells, we have undertaken a 2D-PAGE-based comparative MS-proteomic analysis of this Brd2-transgenic lymphoma in relationship to proliferating and resting B cells. These results confirm and extend our recently published study of Brd2 complexes (G. Denis, et.al. J. Proteome Res. In press), which defined Brd2-associated proteins that regulate transcription.

Method: We used 2D-PAGE separation of murine proteins from extracts of normal resting B cells, normal proliferating B cells and malignant Brd2-transgenic proliferating B cells to define 2D reference maps. Importantly, all cells were derived from syngeneic mice, minimizing genetic variation between individuals. Resolved, stained and quantitated protein spots were excised and digested with trypsin. Peptides were desalted and subjected MALDI-TOF MS and peptide mass fingerprinting analyses or to LC-MS and MS/MS. Spectra were analyzed with MoverZTM or MassLynxTM and ProteinLynxTM software, and peptide mass fingerprinting analysis utilized MASCOT and/or BUPID. Identifications from analyses of cell extracts were compared to those obtained through analyses of samples further purified by Brd2-immunoaffinity chromatography to determine identities of proteins associated with Brd2 transcription complexes.

Preliminary data: We have hypothesized that normal proliferating B cells and malignant proliferating B cells share the induction of important proliferative proteins in comparison to non-mitogenically-stimulated controls, however, that malignant cells show additional induced protein expression and post-translational modifications unique to their malignant state, over and above those changes associated with cell proliferation alone. To explore this hypothesis, we generated preliminary 2D reference maps of extracts from normal resting B cells, normal proliferating B cells and malignant Brd2-transgenic proliferating B cells, all derived from syngeneic mice. We have compared these maps and have quantified many proteins that appear to undergo significant changes in expression between states, and well as proteins that appear to be unique to particular states, or have unique patterns of post-translational modification in particular states. Additionally, we have compared these maps to those made from samples subjected to further purification by Brd2-immunoaffinity chromatography. Ongoing MS analyses are being employed to determine the identity of the proteins that change between states and have so far lead to the positive identification of, among others, several important transcription factors and co-activators, which contribute to transcriptional control in both proliferating and lymphomic B cells. These factors were associated with general transcription factors, histones and histone modification enzymes, as well as nucleosome remodeling complexes. Results from our ongoing study illustrate with detail that cancer is more complex than cellular proliferation, rather that it involves additional expression changes in many proteins, several of which are important for control of chromatin status and activated and basal transcription. Moreover, in shedding light on the protein mechanics particular to proliferation and carcinogenesis, our results may lead to the identification many novel targets for chemotherapeutic intervention.

Acknowledgements: ACS grant RSG-05-072-01, NIH grants P41-RR10888, S10-RR15942, CA075107, CA102889, NHLBI contract N01-HV-28178, and the BU Undergraduate Research Opportunity Program.

A Software Shell for MS Data Processing and Management

Yang Su (1,5), Sequin Huang (1,3,6), Hua Huang (1,3), David H. Perlman (1,3), Mark E. McComb (1,3), Catherine E. Costello (1,2,4)

1) Cardiovascular Proteomics Center, 2) Mass Spectrometry Resource, 3) Department of Medicine, 4) Department of Biochemistry, Boston University School of Medicine, 5) Bioinformatics Program, Boston University, Boston, MA, 6) Present address: Waters Corp, Milford, MA

Novel Aspects: A stand alone GUI based software shell for automated MS data file conversion and database search

Introduction: Here we present a stand-alone software application which is designed to aid the processing and management of various MS data files. For data processing it performs two separate but related functions: automatic conversion of processed MS data files (includes both mzXML and mzData formats) into several common formats accepted by different software applications and the submission of these converted data files to several web based database searching engines (Mascot, Ms-Fit, Aldente, BUPID and X! Tandem). For data management, the software supports multiple users/projects and organizes data and settings individually for each user/project. Finally, all these functions are accessible through a user-friendly graphical interface and the executable program will be available for free download.

Methods: The software was developed using Microsoft Visual Studio 6.0. MSXML 4.0 was used as an XML parser to decode Base64 encoded peak list data in the mzXML/mzData file. A C++ library was also built to improve computational efficiency of decoding. Additional supported data formats are intermediate files converted from raw data files using manufacturers' software: LC MS/MS data is processed with Analyst QS (ABI/Sciex), MassLynx/PLGS2.1 (Waters); MALDI MS data with MoverZ (Proteometrics LLC); and FTMS data with BUDA (BUSM). These files can then be converted into a variety of database searching engine formats (web-based formats include Mascot, Ms-Fit, Aldente, BUPID, X! Tandem and stand-alone formats include PLGS and Sequest).

Results: We tested the program using large volumes of MALDI-TOF MS, MALDI-FT MS and LC MS/MS data sets obtained in house. After initial processing using vendor-specific programs our software converted processed files to the formats of several commercially and publicly available search engines. Files were then submitted for protein identification to the search engines with the search settings specified by the user. Search settings are retained and applied to multiple search sessions once they had been entered for the first search. Results files are automatically saved in HTML format and can then be viewed directly inside the program. Both data conversion and database searching were run in the batch mode to handle large amount of data with little manual intervention (BUPID). The program supports multiple users and for each user stores data/result files in his own directory for easy data access. Each user can also save their usual settings into XML files and such files can be automatically loaded. In addition to multi-user support, our program also implemented project support. A typical project consists of information on specific user, source MS data files, processed data files, database search engine settings and search result files. By saving all these related information about a certain project, our software provides an excellent way of managing and archiving proteomics datasets.

Conclusions: The software provides an easy-to-use graphical interface for performing MS data processing and management in unattended batch mode. It can also be easily be expanded for more MS data types and linked to more database search engines.

Acknowledgement: This project was funded by NIH NHLBI contract N01-HV-28178 and NIH NCRR grants P41-RR10888 and S10-RR15942.

Pattern-based Identification of Protein Structural Variations by Mass Spectrometry

Hua Huang, Mark E. McComb, David H. Perlman, Hongyuan Luo, Timothy P. Skelton,
Martin H. Steinberg, David H. K. Chui, Catherine E. Costello
Boston University School of Medicine, Boston, MA

Novel Aspects: Pattern-based identification for important protein structural changes in biological samples by MALDI-TOF-MS and nanoLC-ESI-MS/MS coupled to an automated proteomics platform.

Introduction: Mass spectrometry is widely used for characterization of protein primary structure and structural changes, such as sequence mutations and post-translational modifications. MALDI-TOF-MS is effective for peptide mass mapping, and LC-MS/MS is a powerful technique in protein/peptide sequencing and in localizing the site of structural changes. For human hemoglobin, many mutations and post- or co-translational modifications have been found associated to biological functions and disease states. In this presentation, we demonstrate the pattern-based identification of peptide structural changes for MALDI-TOF-MS data and LC-MS/MS data. These pattern changes can be in intensity or in isotopic peak distribution, or both. The LC-MS/MS measurements of these samples confirmed their identification based on the MS results.

Methods: Whole blood was diluted and cleaned by centrifugal filtration (Millipore) to remove cellular debris and salts. Trypsin digestion or AspN digestion of intact globin chains was performed for peptide mass mapping and tandem mass spectrometric measurement. The digests were analyzed by MALDI-TOF MS (Bruker Reflex IV) and online nanoLC-MS/MS (QTOF-API-US, Waters Corporation). MALDI-TOF-MS data were analyzed by MoverZTM, and LC-MS and LC-MS/MS data were processed and searched against SwissProt and custom programmed Hemoglobin/PTM databases using commercially available (ProteinLynx Global Server 2.2, Waters Corporation) and software written in-house.

Results: MALDI-TOF-MS has been performed for the tryptic digests of the samples. High sequence coverage of up to 95% for both alpha and beta globin chains was routinely achieved, assuring the identification of any variants and possible post- or co-translational modifications in the globin chains. We have identified variants and modifications by a pattern-based identification methodology either in intensity or in isotopic peak distribution, or both. The structural variations identified include the beta globin chain mutation Asp52 to Asn with a 1 mass unit decrease and the beta globin chain Lys95 mutated to Glu with a 1 mass unit increase. The alpha globin chain mutation Asp47 to His has also been identified as distinct from sodium adduct with an isobaric 22 mass unit increase. Diverse hemoglobin variants have been identified consistent with their DNA sequencing results. Additionally, post-translational modifications have also been observed in these biological samples. The MS/MS sequencing of the peptides containing the structural changes has been achieved by nanoLC-MS and MS/MS, and results confirmed and extended the pattern-based identifications by MALDI-TOF-MS. By applying automated data processing and databank searching against a pre-programmed hemoglobin database and a pre-programmed post-translational modification database, the large LC-MS/MS data sets could be analyzed in a high-throughput mode. This complementary methodology combining pattern-based identification for MALDI-TOF-MS data with LC-MS/MS sequencing information has demonstrated its power for characterizing subtle primary structural changes in proteins and peptides.

Acknowledgement: This project was funded by NIH grant P41 RR10888 and NHLBI contract N01 HV28178. We are grateful for software assistance from Bruker Daltonics and Waters Corporation.

A Software Application for MS Data Processing and Post Translational Modification Database Generation

James West, Hua Huang, Weiwei Tong, Yang Su, David H. Perlman, Mark E. McComb, Catherine E. Costello
Cardiovascular Proteomics Center, Boston University School of Medicine, Boston, MA

Novel Aspects: Use of nominal and exact mass PTM databases for MS analyses of PTMs

Introduction: MS based proteomics is only just beginning to be applied to the characterization of post translational modifications (PTMs) of proteins. By this means, with accurate mass tandem mass spectrometry (MS/MS) data obtained from LC-MS/MS experiments of proteolytic digests of proteins it is possible to characterize a limited number of known PTMs. However, for the discovery of unknown PTMs or the identification of unknown localization sites, the use of a standard PTM database fails. We have constructed a database of nominal mass and exact mass PTMs and have applied a series of successive algorithms to determine the feasibility of using this approach for PTM characterization.

Methods: The PTM Database Generator (PTM-DBGen) software application was developed using Microsoft Visual C# .NET utilizing the Microsoft .NET Framework 1.1 (version 1.1.4322 SP1) in the Microsoft Visual Studio .NET 2003 Development Environment and was used to generate custom nominal and exact mass PTM databases in XML. Tryptic digests of proteins from several in-house projects were analyzed by online nanoLC-MS/MS (QTOF-API-US, Waters Corporation). The LC-MS and LC-MS/MS data were processed using ProteinLynx Global Server 2.2 (Waters Corporation), which is commercially available, and searched against the SwissProt database and the custom generated protein and PTM databases produced with PTM-DBGen. PTM Database subsets were also generated from the RESID database using PTM-DBGen.

Results: The nominal and exact mass PTM databases generated using PTM-DBGen were tested using large volumes of existing LC MS/MS data sets obtained in house. Initial results were successful, however several anomalies were detected: 1) Isobaric PTMs tended to generate erroneous results and false positives, 2) Amino acid substitutions and/or sequence homology may be misidentified as potential PTMs, 3) Mass shifts from the nominal mass resulted in increasingly large mass errors (in ppm), which may result in the removal of a potential PTM result from the database search, 4) while searching for 2 PTMs was feasible, and resulted in the identification of multiple PTM sites within a single peptide, several false positives were observed as well. To minimize the number of false positives a sequential and iterative approach was used. Data were first searched against a non-redundant database. After this first pass all positively identified peptides were removed from subsequent searches and the resulting protein list used to create a secondary protein database. Secondary searches were then directed against this database and positively identified peptides removed from the data set. An amino acid variant and sequence homology search was then conducted followed by the first and subsequent PTM database searches. This approach minimized false positives and afforded the return of known and novel PTMs.

Conclusions: The use of accurate mass, a sequential and iterative search algorithm and strict positive ID settings for the database searches resulted in the minimization of false positives and allowed for the identification of several PTMs based on a nominal and exact mass database.

Acknowledgement: This project was funded by NIH NHLBI contract N01-HV-28178 and NIH NCRR grants P41-RR10888 and S10-RR15942.

**Intact Protein Separation and Identification by Non-porous
Reversed Phase Liquid Chromatography and Mass Spectrometry**

Claire Dauly (1), David H. Perlman (1), Hua Huang (1), Mark E. McComb (1), Catherine E. Costello (1,2)
1) Cardiovascular Proteomics Center, 2) Mass Spectrometry Resource
Boston University School of Medicine, Boston, MA.

Novel Aspects: Intact protein and peptide mapping analysis by reversed phase HPLC MALDI-TOF MS and comparison to SDS-PAGE electrophoresis.

Introduction: Protein fractionation is a challenging goal in proteomics to decrease sample complexity prior to MS analyses. SDS-PAGE gel electrophoresis is a commonly used technique to separate proteins but remains difficult to automate and suffers from a lack of reproducibility and MS sensitivity. It is also very difficult to obtain intact protein MS information from SDS-PAGE gels. Here non-porous reversed phase HPLC coupled with MS was investigated as an alternative to SDS-PAGE methodology. Fractionated proteins can thus be directly analyzed by MALDI-TOF MS or digested for identification via peptide mass database searches. Results were compared to SDS-PAGE gel resolution, reproducibility, Coomassie stain sensitivity and to peptide recovery after tryptic digestion. As an application, a membrane protein extract was characterized.

Methods: A mixture of proteins with a broad range of pI and molecular masses was prepared for use as a standard to characterize the HPLC fractionation system. Each protein was run individually and as a mixture through the np-RP column to investigate the sensitivity of UV-detection, the reproducibility, the linearity and the capacity of the system. Protein fractions were then analyzed directly by MALDI TOF-MS for molecular weight information or digested with trypsin and the resulting peptides analyzed by MALDI-TOF MS followed by database searching and identification. Results were compared to SDS-PAGE separation followed by in-gel digestion in terms of peptide recovery and sequence coverage. A membrane protein extract was then fractionated on the np-RP column and then characterized by MS.

Results: The np-RP column was shown to be very reproducible in term of protein separation and UV detection and to exhibit an excellent linearity of detection. These two proprieties are crucial for proteome profiling and differential proteomics analysis. The detection limit of the UV detector was estimated in the ng order. Intact proteins were analyzed by MALDI-TOF MS for molecular weight estimation. Sensitivity of 50 pg (10 fmol) of intact protein was achieved by MS after the np-RP column separation. Protein fractions were then digested by trypsin following a protocol that we developed for this project and the resulting peptides were analyzed by MALDI-TOF MS. The integrated HPLC MALDI-TOF-MS system was shown to be as efficient as SDS-PAGE separation and in-gel digestion in terms of peptide recovery and sequence coverage. A membrane protein sample was used as an application of this proteomic methodology. The intact masses of the proteins present in each fraction were assessed and confirmed the results of protein identification obtained from peptide mass fingerprinting analysis of the digested fractions. Among the identified proteins, vimentin was observed due to its abundance in the sample. Vimentin, a regulated intermediate filament protein, is known to interact strongly with the plasma membrane.

Overall we demonstrated that the HPLC MALDI-TOF MS system allows proteins from a complex sample to be rapidly fractionated and identified by integration of the intact molecular weight information and the peptide mass database search. Using this methodology, intact proteins may also be easily analyzed directly by MS, a feat which remains difficult when using SDS-PAGE electrophoresis.

Acknowledgements: This project was funded by NIH grants P41 RR10888 and S10 RR15942 and NHLBI contract N01 HV28178. We are grateful for technological support from Beckman Coulter Corporation.

**Structural Characterization of Glycosphingolipids and Toxin Receptor Gangliosides
by IRMPD With TLC/VC-FTMS**

Vera B. Ivleva¹; Anne A. Wolf²; Wayne I. Lencer²; Peter B. O'Connor¹; Catherine E. Costello¹

¹Boston University School of Medicine, Boston, MA; ²Children's Hospital, Boston, MA

Novel Aspect: The ceramide structures of the glycosphingolipids and gangliosides desorbed directly off the TLC plate were analyzed by SORI-CAD and IRMPD-FTMS.

Introduction

Glycosphingolipids and gangliosides participate in diverse biological processes, and their biological roles are dependent on the structures of both the oligosaccharide and the ceramide portions. Here, we use vibrationally cooled MALDI-FTMS for the detection of labile species followed by their efficient fragmentation by SORI-CAD and IRMPD. GM1 and GD1a gangliosides serve as trafficking receptors for cholera toxin and the related LTIIb toxin, respectively (1). Structural variation in the lipid portion of the ganglioside GD1a is being evaluated as a moderator of this function.

Methods

The glycosphingolipids were dissolved in CHCl₃/MeOH and applied onto the MALDI target with sinapinic acid as a matrix. For TLC-MALDI FTMS analysis of the gangliosides, the plate was sprayed with the saturated matrix solution and affixed to the target. Fragmentation was performed by SORI-CAD with N₂ collision gas and IRMPD techniques. We have previously described ganglioside separations and instrumental parameters for VC MALDI-FTMS (2). In the current study, the samples are MALDI-desorbed directly off TLC plate surfaces with ~0.5 mm sampling steps and thermalized by the cooling gas. The ions are then subjected to dissociation. We are using polarized intestinal epithelial cell line T-84 and monkey kidney Vero cells for ganglioside purification (1).

Preliminary Results

The preliminary results showed that IRMPD of the glycolipids results in efficient fragmentation of the oligosaccharide and the ceramide moieties, whereas SORI-CAD requires multi-step fragmentation with an optimization of the SORI pulse for each individual fragment. The fatty acid and sphingosine fragments of the glycolipids with larger glycans could only be obtained by the IRMPD technique. In the positive ion mode, the predominant peaks among the ceramide fragments are those containing the unsaturated sphingosine chain. Characteristic fatty acid fragment ions were observed in the negative ion mode as well. The fragmentation was obtained for the species desorbed directly from the plate after TLC separation. Compared to glycosphingolipids, the ganglioside molecules required higher energy deposition for efficient fragmentation. An abundant diagnostic ion resulting from ceramide cleavage was produced by the gangliosides in the negative ion mode. Lithiated samples fragmented more readily, but lithium deposition should be avoided for the TLC-separated samples to avoid dilution. The IRMPD fragmentation method is being applied to the analysis of the toxin receptor gangliosides after their TLC separation. The preliminary MS-scanning of these samples showed a significant heterogeneity of the ceramide and glycan structure.

Acknowledgement: The project has been funded in part with Federal funds from the National Institutes of Health, under grant No. P41 RR10888 from the National Center for Research Resources and contract No. N01-HV-28178 from the National Heart, Lung, and Blood Institute to CEC; NIH DK48106 and DK57827 and DK34854 to WIL and DK02934 and an Investigator Development award from the CDHNF to AW.

1) Lencer, W.I. *Int. J. Med. Microbiol.* 2004, 293(7-8), 491-494.

2) Ivleva V.B., Elkin, Y.N., Budnik B.A., Moyer, S.C., O'Connor, P.B., Costello, C.E. *Anal. Chem.* 2004, 76, 6484-6512.

Keywords: Chromatography, Thin layer; Fourier Transform ICR; Glycolipids; IRMPD; SORI (sustained off-resonance excitation)

A Detailed Map of Oxidative Post-translational Modifications of Human p21ras using Fourier Transform Mass Spectrometry

Cheng Zhao; Mahadevan Sethuraman; Nicolas Clavreul; Parminder Kaur;
Richard Cohen; Peter O'Connor
Boston University School of Medicine, Boston, MA

Novel Aspect: The study provides an effective and efficient method to map and predict relative selectivity and specificity of oxidative post-translational modifications.

Introduction: P21ras, the translation product of the most commonly mutated oncogene known, is a small guanine nucleotide exchange protein. Oxidant-induced post-translational modifications of p21ras including *S*-nitrosation and *S*-glutathiolation have been demonstrated to modulate its activity. Structural characterization of this protein is critical to further understanding the biological functions of p21ras. For peroxynitrite-treated p21ras, five oxidized methionines, five nitrated tyrosines, and at least two oxidized cysteines (including C118) were identified by “bottom-up” analysis with high resolution and high mass accuracy FTMS and the major oxidative modification of C118, Cys¹¹⁸-SO₃H, was confirmed by several tandem mass spectrometry experiments. Additionally, “top-down” analysis was conducted on p21ras *S*-glutathiolated by oxidized glutathione, and, identified C118 as the major site of glutathiolation among the four surface cysteines.

Methods: In this study, FTMS was utilized to map, in detail, the post-translational modifications of p21ras exposed to oxidants by combining bottom-up and top-down techniques. Recombinant p21ras (10-20 µg) was dialyzed to remove commercial buffers components and treated for 5 min at 37°C with peroxynitrite (ONOO⁻, 100-250 µM) or with GSSG (100 µM) in phosphate buffer (0.1 M, pH 7.0). For “bottom-up” analysis, the control and the modified samples were digested with trypsin at 1:25 (w/w) in 100 mM ammonium bicarbonate (pH 8.0) for 3 h at 37°C and purified with Poros 50 R1 material. For “top-down” analysis, the control and the modified samples were purified with Poros 50 R1 material.

Preliminary data:

Peroxynitrite-treated p21ras: Due to the complex and heterogeneous oxidative modifications on p21ras, the top-down analysis proved difficult to perform. Thus, the bottom-up method was utilized to map the oxidative post-translational modifications on this sample. Fortunately, 100% sequence coverage was obtained. After internal calibration, most of the errors between experimental and theoretical mass of these peptides are within 2 ppm. Five methionines (M1, M67, M72, M111, M182) were oxidized and five tyrosines (Y4, Y40, Y96, Y137, Y157) were nitrated. Cys118, which resides in the GTP binding site, was oxidized. Cys118 sulfenic acid, sulfinic acid, sulfonic acid, and *S*-nitrosothiol were all observed though those with sulfenic acid, sulfinic acid and nitrosothiol modifications were very low abundance. At least one of the three terminal cysteines (Cys 181, Cys184, and Cys186) was oxidized into Cys-SO₃H. Peaks corresponding to two and three terminal cysteines with sulfonic modifications were also detected, but they were much lower intensity than the one with only one sulfonic modification. The methionine and cysteine modifications on peptide 103-123 were confirmed by both low energy CAD and ECD.

S-Glutathiolated p21ras. At least three groups of charge state distributions were observed in *S*-glutathiolated p21ras ESI spectrum and the most intensive one indicated one glutathione addition and the other two indicated two and three glutathione additions compared to the unmodified p21ras. Bottom-up analysis showed all the cysteines were glutathiolated, although some of them maybe formed by exchange of disulfide bonds between GSSG and p21ras protein/tryptic peptides. It is impossible to determine which cysteine is the major site of glutathiolation, this difficulty was solved with top-down analysis of the major glutathiolated p21ras. Thirty-two cleavages in CAD, 46 cleavages in ECD, and 4 complementary pairs (100% sequence coverage) were obtained in the top-down analysis and C118 was revealed to be the major glutathiolated cysteine.

Acknowledgement: Supported by NIH P41-RR10888, N01-HV28178, R01 AG27080, P01 HL HL81587, MDS Sciex, the ACS Petroleum Research Fund, and an NIH Post-doctoral Research Fellowship T32 HL07224.

Key words: Dissociation (CAD), Electron Capture Dissociation, Mass spectrometry Fourier Transform, Oxidation, Post-translational modification

Probing the Long-lived Radical Intermediates in Electron Capture Dissociation with a Double Resonance Experiment

Cheng Lin; Jason J. Cournoyer; Peter B O'Connor
Boston University School of Medicine, Boston, MA

Novel Aspect: Double resonance experiment revealed the direct correlation of the ECD product ions with a long-lived radical intermediate.

Introduction: Despite its wide application in the field of proteomics, the mechanism of electron capture dissociation (ECD) is still not thoroughly understood. Recent calculations and experiments suggested that in addition to a rapid, possibly non-ergodic primary cleavage, ECD generates a metastable radical cation as one of the product species. ECD of cyclic peptides showed extensive secondary sidechain and backbone cleavages, which was postulated to be the result of free radical cascade. Radical rearrangement was also observed in linear peptide ECD in the form of H/D scrambling. The presence of reaction intermediates can sometimes be revealed via double resonance (DR) experiments, where abundance drop of one ion upon resonant ejection of another indicates that the former forms through the latter.

Methods: Experiments were carried out on a home-built hybrid Fourier-transform mass spectrometer (FT-MS) with an electrospray ionization (ESI) source. A single charge state of peptide/protein ions of interest were isolated and externally accumulated in the front end quadrupole before entering the cell, where they were trapped and subjected to ~2-10 millisecond irradiation of ~0.5 eV electrons from an indirectly heated dispenser cathode. In the double resonance experiment, the charge reduced molecular ions were ejected simultaneously from the cell by applying single frequency excitation with a peak-to-peak amplitude of 5 to 40 volts during the electron irradiation period. All spectra were zero-filled twice, Fourier transformed without apodization, and internally calibrated with the molecular ion, charge reduced species and their isotopic peaks.

Preliminary Results:

ECD of fibrinopeptide B (EGVNDNEEGFFSAR) generated a complete z-ion series, with the peak intensity of every z-ion smaller than z₉ decreasing sharply upon resonant ejection of the charge reduced molecular ions, while that of z₉ and larger ions remaining roughly the same. It is well known that after the initial N-C bond cleavage, the c/z⁺ ion pair may still be held together by noncovalent interactions. These ion pairs may be ejected out of the cell before they have a chance to break up, causing the drop in the peak intensity of the individual fragment ions detected. In fibrinopeptide B, the C-terminal arginine can form hydrogen bonds with the asparaginy, aspartyl and glutamyl sidechains. It is thus of no surprise that smaller z-ions showed steeper drop in their peak intensities, as their complementary c-ions have more such sidechains available to interact with arginine. Since the ejection time directly relates to the ejection voltage, it is possible to find the lifetime of these intermediates. Two of such ion pairs were determined to have a lifetime of around 2 milliseconds.

Much more surprisingly, fragment ions from ECD of cyclosporin A showed virtually no change in their abundances in the DR-ECD experiment, even though almost all fragment ions are secondary ions. Thus the secondary backbone cleavage via free radical cascade must occur on a timescale shorter than the ejection time, and the resulted secondary fragment ion pairs fell apart readily, due to lack of strong noncovalent interactions of this aliphatic residue rich peptide.

Most ECD fragment ions from bovine ubiquitin 7+ ion decreased in their abundances with reduced molecular ion ejection, while many from 11+ ion showed little change, indicating the former to be of a compact form, and the latter being more unfolded in the gas phase.

Acknowledgment: This work was supported by MDS SCIEX, NIH/NCRR P41RR10888, NIH/NHLBI N01HV28178, and the ACS Petroleum Research Fund.

Keywords: Electron-Capture Dissociation; Fourier Transform ICR; Fragmentation Mechanism/Pathways; Protein Conformation

Determination of High-Precision Isotope Ratios from Experimental Isotopic Distributions

Parminder Kaur and Peter B. O'Connor
Boston University, Boston, MA

Novel Aspect: Use of high-precision isotope ratio measurements with any kind of instrument for any given sample

Introduction

Isotope variability due to natural processes provides important information for studying a variety of complex phenomenon like determining the genesis of a given sample, dietary studies of species, nitrification rates in trees etc. These measurements require very high-precision determination of isotope ratios of a particular element involved. Isotope Ratio Mass Spectrometers (IRMS) are widely employed tools for such a high-precision analysis. IRMS instruments accept the sample analyte in the form of only a limited number of gases which must represent the isotopic characteristics of the original sample, which causes lack of flexibility. This work aims at overcoming the limitations inherent to IRMS by estimating the elemental isotopic abundance from the experimental isotopic distribution.

Methods

Experimental isotopic distribution is an indirect measure of the isotopic abundances of various elements present in the molecule. It can be represented by the joint convolution of the isotopic abundances of each of the individual atoms. Mathematical techniques have been developed in order to factor out the known isotopic abundance contributions from various elements followed by calculation of the corresponding unknown values for the element of interest. An estimate for the required isotopic abundance is generated from each of the observed isotopic peaks by solving the convolution equations, and the final result is reported to be the mean of each of the individual values obtained.

Preliminary Results

Computer generated simulations were carried out in order to generate the experimental isotopic distributions for a given molecule with known elemental composition and isotopic abundances. The results thus obtained were subjected to the developed theoretical framework in order to estimate the isotopic abundance for Carbon from each isotopic peak, with the abundance values for the other elements being taken into consideration in the calculation. These estimated results are shown to be in good agreement with their true values. Increasing the number of ions for generating the experimental isotopic distribution greatly improves the estimate. This is because in the limit of infinite number of ions, experimental isotopic distribution approaches its theoretical counterpart and is a true representative of the composition of its constituents. High molecular weight molecules are shown to be particularly advantageous because of the presence of larger number of isotopic peaks in the isotopic distribution leading to a greater amount of information. Initial results reveal that with sufficiently high number of ions and multiple experiments, it is possible to distinguish between the samples varying very slightly in the Carbon isotopic abundance. For example, the estimate can distinguish whether the sample originated from marine plankton (C13 abundance = 1.09%, Delta C13=-19.5) or meat from an animal feeding on a C4 plant (C13 abundance=1.1%, Delta C13=-12.5). Results from the experimental data will be presented comparing the true abundance value with its estimated counterpart. This approach eliminates the need to convert the sample into gaseous form. The results are applicable to any type of mass spectrometer, and for any type of sample. The results can also be extended to estimate the isotopic abundance of any unknown element provided the isotopic abundance of the other elements are known *a priori*.

Acknowledgment: This work was supported by NIH/NCRR P41RR10888, NIH/NHLBI N01HV28178, and the ACS Petroleum Research Fund.

Keywords: Computational Methods; Data Analysis; Fourier Transform ICR; Isotope; Isotope Ratio MS

Electronics for the Cryogenic Fourier Transform Mass Spectrometer

Raman Mathur, R.W. Knepper, Peter B. O'Connor

Novel Aspect: New electronics for the Fourier Transform Mass Spectrometer.

Introduction

In an FTMS, application of precise voltages and currents is essential for successful transfer and detection of ions in the ICR-cell. Here we report the development of the necessary electronic components for an FTMS, including:

1. Detection Preamplifier – A differential amplification circuit using GaAs devices has been designed for operation at 4Kelvin.
2. Printed Circuit Board (PCB) for R.F. power generator – PCBs of an R.F. generator to drive multipole ion guides have been constructed and tested on ESI-FTMS.
3. Isolation PCBs for Analog and Digital channels – A PCB is designed to protect the channels of the analog output board from high voltage/current transients. Another PCB is designed to prevent crosstalk on the digital output channels.

Methods

The designed preamplifier is mounted on a specially designed PCB for in-vacuum operation close to the detection plates of the ICR cell. The PCB is thermally attached to the 4 Kelvin tube to reduce the thermal noise in the components of the preamplifier circuit. The preamplifier has been tested on bench using a high frequency spectrum analyzer for stability and gain.

The 2 layer PCBs for the R.F. power generator occupies 178 x 128 x 128 mm³ including the transformer coil. Special design constraints were needed for High voltage R.F. on the PCB.

Results

The current preamplifier design has achieved a differential voltage gain of 10 with a 3 dB bandwidth of 1.45 MHz. This design should allow about a ten-fold improvement in signal/noise ratio compared to the current designs at room temperature, and has been created using components that are compatible to cryogenic temperature operations, which will potentially reduce the noise another 8 fold. The operating point of the FETs is set to $V_{DS} = 1.5$ volts and $I_{DS} = 5$ mA which has limited the power dissipation in the preamplifier circuit to be less than allotted 100 mW for the complete circuit. The functionality of the preamplifier was tested in a 7 Tesla magnetic field with the channel of the FET at various angles to the magnetic field. The operating point of the FETs varied slightly, however the gain-bandwidth of the preamplifier remain unchanged. A cryostat has been designed to test the amplifier at 4 Kelvin.

The R.F. power generator PCB was used to drive the hexapole of a home-built ESI-FTMS with a voltage signal of 300 volts p-p and 800 KHz. Ubiquitin ions with charge states +7 to +14 were successfully transferred and detected in the ICR cell.

Isolation PCBs using Channel protectors and Transient Voltage Suppressors were also designed to prevent any damage to the PXI-based analog and digital output cards.

Acknowledgements: Authors would like to thank Kostantin Aizikov for the help in construction of transformer coils for R.F. oscillator. This work was sponsored by NIH/NCRR P41RR10888 and by NIH/NHLBI N01HV28178.