Published in final edited form as:

J Mass Spectrom. 2013 March ; 48(3): 340–343. doi:10.1002/jms.3164.

Mass biased partitioning to enhance middle down proteomics analysis

Joe R. Cannon¹, Nathan J. Edwards², and Catherine Fenselau¹

¹Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742

²Department of Biochemistry and Molecular & Cellular Biology, Georgetown University Medical Center, Washington, DC 20057

Abstract

A strategy is presented for enhancing the middle-down analysis of higher mass peptides recovered from complex protein mixtures. Following a 30 min digestion of multiple myeloma cell lysate by an acid cleavage reaction that is selective for aspartic acid, a 3000 Da membrane filter is used to bifurcate the peptide product mixture, and the heavier fraction is subjected to collisional activation with precursor selection that excludes charge states below +4. Filtration and charge state selection are shown to provide significant increases in the number of peptides identified in the mass range above 3000 Da and in information about protein sequences.

Introduction

Middle-down proteomic strategies have been designed to exploit the advantages of analyzing heavier peptides (3000–20,000 Da) in proteomic analyses.^{1, 2} These include improved chromatographic fractionation, higher sequence coverage, and characterization of cohabiting and potentially interactive modifications.^{3, 4} Experimentally, analysis of peptides in the mass range 3000 to 20,000 Da simplifies the complex mixtures offered by bottom up strategies, while avoiding the diminished performance of top down experiments.^{5, 6} Most often, single-residue specific enzymatic reactions are used since they produce a comparatively higher proportion of mid-range peptides than the standard tryptic digestion. Previously we have demonstrated the use of a chemical method that cleaves proteins selectively at aspartic acid in less than 30 min.⁷ Like other single residue proteolytic agents, microwave accelerated acid cleavage of complex protein mixtures produces complex peptide mixtures that contain enhanced numbers of heavier peptides (>3000), and still a majority of lower mass peptides (<3000Da). During automated analysis by LC-MS/MS, the abundance of low mass peptides in the mixture can suppress or obscure precursor selection and activation of the mid-range peptides. In addition, higher mass peptides require higher mass resolution to allow product ion spectra to be deconvoluted to extract ion masses. For this and other reasons, the optimal conditions for analysis of bottom up and middle down sized peptides are different and cannot be employed simultaneously. Here we report the use of molecular weight cutoff filters to separate a complex peptide mixture into high mass and low mass fractions. In addition, we limit precursor selection in high resolution tandem mass spectrometry experiments to charge states of 4+ or greater in order to optimize high throughput analysis of mid-range peptides. Sample preparation combining these two experimental modifications is evaluated here as mass biased partitioning.

Materials and Methods

RPMI 8226 Cell Culture, Isolation, and Lysis

RPMI 8226 multiple myeloma cells were grown and harvested as published.⁸ The cells were lysed in 2.5 mM imidazole, pH 7, using nitrogen cavitation at 1250 psi (Parr Instrument Co, Moline IL). Following centrifugation at 10,000 rpm for 10 min, the supernatant was collected and stored at -80° C until digestion.

Microwave Supported Acid Hydrolysis and Mass Biased Partitioning

Protein was precipitated from solution using the method of Wessel and Flugge.⁹ The pellet was redissolved in buffer and the concentration was determined by Lowry assay (BioRad, Hercules, CA). A 50 µg aliquot was diluted to 0.1 µg/µL with acetic acid and water to arrive at a final solvent concentration of 12.5% (v/v) acetic acid. The acidified sample was digested in a CEM Discover Microwave (Matthews, NC) at a maximum temperature of 140° C for 30 minutes while irradiating with 300 W. Following digestion the samples were allowed to cool to room temperature. Amicon 3 kDa and 10 kDa molecular weight cutoff filters (Millipore, Billerica, MA) were equilibrated with 12.5% acetic acid prior to use. The acid digestion products were then fractionated through the filter according to the manufacturer's instructions. The high mass fraction retained above the filter was diluted with 200 µL of Milli-Q water, aspirated 30× to maximize recovery, and lyophilized to reduce the volume to approximately 100 µL. The low mass filtrate was also collected and concentrated to 100 µL.

LC-MS/MS

Due to the higher abundance of low mass peptides for the low mass sample, 1/20th of the total volume was injected for LC-MS/MS analysis. Reversed phase chromatography was carried out using a Shimadzu Prominence LC (Columbia, MD).

Following injection, the peptides were desalted and concentrated online for 20 minutes at a flow rate of 10 µL/min with 100% Solvent A (97.5/2.5/0.1 H₂O/ACN/formic acid). The peptides were then fractionated using a 0.150 mm × 150 mm Grace Vydac Everest C18 column packed with 5 µm particles with 300 Å pores (Deerfield, IL). The flow rate was set to 300 nL/min and the concentration of Solvent B (97.5/2.5/0.1 ACN/H₂O /formic acid) was increased in a linear fashion from zero to 35% over the course of 180 min. Low mass peptides were introduced into an LTQ-Orbitrap XL and MS1 scans were acquired at 30,000 resolving power. Precursor peaks were limited to the 8 most abundant multiply charged peptides. CID product ion spectra were recorded in the LTQ at unit resolution.

For the high mass portion, 100 µL of sample was injected for each analysis in the LC-LTQ-orbitrap. Chromatographic conditions were identical to those outlined above. Survey scans were acquired at 30,000 resolving power and the three most abundant multiply charged precursors were isolated and fragmented by CID in the ion trap, and subsequently detected at 7,500 resolving power in the orbitrap.

Charge State Selection

For the low mass fraction, data dependent analysis was set to isolate and fragment precursor ions with charge states of +2 or +3. For higher mass peptide analysis, only precursors with charge states greater than +4 were selected for isolation and fragmentation. Automated gain control targets were set to 5×10^5 and 5×10^6 for the survey and product ion scans, respectively. Precursor isolation width was set to 10 Da. This relatively wide window provides higher intensity (high resolution) product ion spectra and thus more peptide

identifications. Incorrect identifications resulting from overlaps are controlled by setting a stringent confidence threshold.⁵

Bioinformatics

Mid-mass peptides were identified using ProSightPC2.0, which incorporates the Xtract algorithm for precursor and product ion neutral mass calculations (ThermoFisher, San Jose CA). Precursor mass tolerance using ProSightPC was set to 250 Da and searched in DeltaM mode. DeltaM mode facilitates identification in the presence of an unforeseen deviation from the candidate peptide mass (such as a mass change of ± 115 Da indicative of the loss or addition of Asp). All N-*or* C-terminal Asp-cleavage products observed in microwave supported acid hydrolysis are easily accommodated using a 250 Da search window in conjunction with DeltaM mode. In addition, the large window allows for the identification of dehydrated and oxidized species. Fragment mass tolerance was set to 15 ppm. Mid-mass spectra were *in silico* searched against a database of reviewed Human proteins from the UniProtKB digested using ProSight's Database Manager utility, with a maximum of 9 missed cleavages and maximum peptide mass of 20 kDa. Spectra were also searched against a shuffled version of the same database. Identifications were assigned automatically with E-values less than or equal to 10^{-8} , the threshold at which no peptides were matched in the shuffled database search. Other peptides from the high mass fraction that were matched with E values $< 10^{-4}$ were validated manually. RAWmeat (Vast Scientific, Cambridge, MA) was used to obtain scan and charge statistics from data files.

The low mass peptide fraction was searched using the PepArML meta search engine, a unified interface for combining results from up to seven different search engines.¹⁰ PepArML combined results from 6 different search engines against the IPI Human database incorporating cleavage specificity at either side of Asp residues and estimated peptide specific false discovery rates (FDR). Peptide identifications were limited to those with FDRs of 10% or less. Dehydration at Asp, pyro-glutamic acid formation from N terminal Glu and Gln, and oxidation of Met were added as variable modifications. For all six search engines precursor mass tolerance was set to 10 ppm. *In silico* digestions were performed using software developed in house.

Results and Discussion

In order to prepare a sample enriched in mid-range peptides, fractionation of the peptide mixture was evaluated using both 3 kDa and 10 kDa filters. The fraction of heavy peptides recovered from each filter was analyzed by LC-MS/MS, excluding selection of +2 and +3 precursor ions and implementing 7500 resolution for product ion analysis as described in the Experimental section. Increasing the resolution for product ion spectra improves product ion charge-state determination and boosts the specificity of product ion matching, however, at the expense of increased acquisition time. A larger proportion of the peptides recovered from the 10 kDa filter were heavy (> 3 kDa), compared to peptides from the 3 kDa filter, however the absolute number of heavy peptides was greater for the 3 kDa filter. Consequently the 3 kDa filter was selected for the rest of the study.

Figure 1 shows the molecular masses of peptides identified from injections of the heavier retentate and the lighter filtrate. In this figure, the proportion of peptides identified in each fraction is plotted side-by-side in each of 500 Da bins. The peptides identified from each injection display distinct mass distributions, reflecting successful fractionation of light and heavy masses. Of the total, 6% of the peptides identified have masses > 3000 Da. To some extent, this reflects the smaller number of heavier peptide products that can be formed. However, a theoretical analysis of Asp C cleavage of the proteins identified indicates that 27% of the peptide products would be expected to weigh more than 3000 Da (with 0 missed

cleavages and a maximum peptide mass of 20 kDa). By contrast, 91% of the peptides identified had masses below 3000 Da, while 73% were predicted. The experimental under-sampling observed in our analysis of the mid-range peptides is due in large part to the low duty cycle of the instrument used. Each cycle of three tandem analyses at 30k/7.5k (see Experimental) required 0.875 sec, while the duty cycle for analysis of low mass peptides, eight tandem analyses at 30k/unit resolution, was 0.5 sec. Chromatography and ion activation also limit the identification of larger peptide ions.

A combined total of 349 proteins were identified from 624 peptides when both the heavy and light peptide fractions were analyzed using optimal conditions for each (summarized in Supplementary Table 1). This ratio of peptides to proteins is consistent with other middle-down studies where single residue proteolysis was used, and large and small peptides were analyzed without size-based fractionation.^{5, 11} Five hundred and sixty-eight peptides were observed from 333 proteins in the low mass fraction, whereas 56 peptides were identified in the high mass fraction (3kDa filter) from 38 proteins. The distribution of experimental peptide identifications was also compared to those from two control experiments. Whole cell lysate was subjected to acid cleavage and analyzed twice, first using the LC-MS/MS parameters optimized for the lighter filtrate and also those optimized for the heavier peptide mixture. One hundred and ninety-eight peptides were identified when product ion characterization was carried out with unit mass resolution in the ion trap (our low mass mode). Fifteen of these peptides had masses exceeding 3000 Da, however the use of low resolution compromised their identification. Twenty-eight peptides were identified using the higher resolution Orbitrap analysis, of which thirteen had masses exceeding 3000 Da. We conclude that mass based fractionation and mass biased analysis provides more peptide identifications overall in whole cell lysate, and, more to the point, a nearly three-fold increase in identifications of mid-range peptides with high reliability.

We have reported previously, in a study of human ribosomal proteins, that analysis of heavier peptides obtained using a single residue cleavage (Asp-selective acid cleavage) provides higher sequence coverage than the shorter peptide products of tryptic proteolysis.¹¹ The advantageous coverage offered by longer peptides is confirmed in the present study of whole cell lysate. Thirty-three per cent of singlet peptide identifications in the high mass fraction provided greater than 10% sequence coverage of their parent proteins, while only 10% of the individual peptides identified in the low mass fraction provided greater than 10% coverage. Overall, the average protein coverage for the high mass peptide fraction is 9.4%, and for the low mass peptides is 4.6% (See Supplementary Table 2). This trend for Asp-selective acid cleavage is in agreement with the general view of single-residue middle-out analysis of proteins.^{1, 3, 5, 6}

The proportion of peptides produced by a proteolytic reaction that fall into the middle-down range of 3 to 20 kDa depends on the specificity of the reagent, the frequency of the targeted residue(s) and experimental control of missed cleavages. Of particular interest is the distinction between the products of trypsin, which cleaves at two amino acid residues--lysine and arginine, and methods that cleave at a single amino acid. Karger and co-workers have suggested that the false discovery rate decreases with larger peptides.³ Goodlett and co-workers have examined the notion that precursor ion densities are not distributed uniformly across the *m/z* landscape in their gas phase fractionation experiments.¹² To explore the density of overlapping peptide masses in our experiment, human proteins in the UniProtKB were digested *in silico* based on cleavage with trypsin, and also on cleavage at the C-terminal side of aspartic acid, and duplicate peptide sequences were removed. For each *in silico* peptide, the number of additional peptides with masses window 10 ppm were counted and tabulated in various non-specificity bins. (A 10 ppm window was used experimentally for low mass fraction precursor ion searches.) The results are summarized in Figure 2, where

it can be seen that 70% of all Asp-C peptides in the mass range 500 to 1000 share their 10 ppm window with 5 or fewer additional peptides, while 19% of the tryptic peptides 500 to 1000 Da have this advantage. The 70% precursor specificity is achieved with tryptic products only at the molecular mass range of 3500-4000 Da. Consideration of Figure 2 indicates that the molecular masses of peptides in the middle-out mass range are more unique than those of their bottom up counterparts. Theoretically, Asp-specific proteolysis (and likely any single residue proteolysis) produces a higher proportion of peptides with nearly unique masses than trypsin. This allows more reliable identifications with lower false discovery rates due to the lower number of searchable peptide candidates within a given precursor mass tolerance.

In summary, we have demonstrated that mass biased partitioning, comprising both membrane cut-off filtering and precursor charge state selection, provides enhanced access to and identification of mid-sized peptides. It also provides improved bottom up analyses of lower mass peptides. Care must be taken with the analysis of the larger peptides. In our experience there is often just one identified per protein. The use of high resolution in the determination of both precursor and fragment ion masses, and the reduced probability that higher masses overlap provides additional specificity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Dr. Yan Wang, Director of the Proteomics Core Facility of the Maryland Pathogen Research Institute for discussion and advice. The work was supported by a grant from the National Institutes of Health, GM021248.

References

1. Boyne MT, Garcia BA, Li M, Zamdborg L, Wenger CD, Babai S, Kelleher NL. Tandem Mass Spectrometry with Ultrahigh Mass Accuracy Clarifies Peptide Identification by Database Retrieval. *J. Prot. Res.* 2009; 8:374.
2. Meyer B, Papasotiriou DG, Karas M. 100% sequence coverage: a modern form of surrealism in proteomics. *Amino Acids.* 2011; 41:291. [PubMed: 20625782]
3. Wu SL, Kim J, Hancock WS, Karger B. Extended Range Proteomic Analysis (ERPA): a new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications- comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR). *J. Prot. Res.* 2005; 4:1155.
4. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequencing analysis by electron transfer dissociation mass spectrometry. *Proc. Nat. Acad. Sci. U.S.A.* 2004; 101:9528.
5. Wu C, Tran JC, Zamdborg L, Durbin KR, Li M, Ahlf DR, Early BP, Thomas PM, Sweedler JV, Kelleher NL. A protease for 'middle-down' proteomics. *Nat. Methods.* 2012; 9:822. [PubMed: 22706673]
6. Taouatas N, Heck AJR, Mohammed S. Evaluation of Metalloendopeptidase Lys-N Protease Performance under Different Sample Handling Conditions. *J. Prot.Res.* 2010; 9:4282.
7. Fenselau C, Laine O, Swatkoski S. Microwave assisted acid cleavage for denaturation and proteolysis of intact human adenovirus. *Internat. J. Mass Spectrom.* 2011; 301:7.
8. Rahbar AM, Fenselau C. Integration of Jacobson's Pellicle Method into Proteomic Strategies for Plasma Membrane Proteins. *J. Prot. Res.* 2004; 3:1267.
9. Wessel D, Flugge UI. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 1984; 138:141. [PubMed: 6731838]

10. Edwards N, Wu X, Tseng C-W. An Unsupervised, Model-Free, Machine-Learning Combiner for Peptide Identifications from Tandem Mass Spectra. *Clinical Prot.* 2009; 5:23.
11. Cannon J, Lohnes K, Wynne C, Wang Y, Edwards N, Fenselau C. High-throughput middle-down analysis using an orbitrap. *J. Prot. Res.* 2010; 9:3886.
12. Yi EC, Marelli M, Lee H, Purvine SO, Aebersold R, Aitchison JD, Goodlett DR. Approaching complete peroxisome characterization by gas-phase fractionation. *Electrophoresis.* 2002; 23:3205. [PubMed: 12298092]

Figure 1. Comparison of peptide mass distributions observed in the low (blue) and high (red) mass fractions following fractionation using mass biased partitioning.

Figure 2.

The proportion of peptides with masses that overlap within 10ppm, produced by *in silico* digestion of a subset of human proteins in UniProtKB using trypsin (top) and Asp C (bottom). Portions in blue indicate regions with 0-5 additional peptides, regions in red have 6-10, in green have 11-20, and in yellow have greater than 20.