Published in final edited form as:

Anal Chem. 2008 July 1; 80(13): 4994–5001. doi:10.1021/ac800254z.

Top-Down Identification and Quantification of Stable Isotope Labeled Proteins from *A. flavus* Using Online nano-Flow Reversed Phase Liquid Chromatography Coupled to a LTQ-FT-ICR Mass Spectrometer

Timothy S. Collier¹, Adam M. Hawkrigde¹, D. Ryan Georgianna², Gary A. Payne², and David C. Muddiman^{1,*}

¹W.M. Keck FT-ICR Mass Spectrometry Laboratory, Department of Chemistry North Carolina State University, Raleigh, North Carolina 27695

²Department of Plant Pathology North Carolina State University, Raleigh, North Carolina 27695

Abstract

Online liquid chromatography – mass spectrometric (LC-MS) analysis of intact proteins (i.e. top-down proteomics) is a growing area of research in the mass spectrometry community. A major advantage of top-down MS characterization of proteins is that the information of the intact protein is retained over the vastly more common bottom-up approach that uses protease-generated peptides to search genomic databases for protein identification. Concurrent to the emergence of top-down MS characterization of proteins has been the development and implementation of the stable isotope labeling of amino acids in cell culture (SILAC) method for relative quantification of proteins by LC-MS. Herein we describe the qualitative and quantitative top-down characterization of proteins derived from SILAC-labeled *Aspergillus flavus* using nano-flow reverse phase liquid chromatography directly coupled to a linear ion trap Fourier transform ion cyclotron resonance mass spectrometer (nLC-LTQ-FT-ICR-MS). *Aspergillus flavus* is a toxic filamentous fungus that significantly impacts the agricultural economy and human health. SILAC labeling improved the confidence of protein identification and we observed 1,318 unique protein masses corresponding to 659 SILAC pairs, of which 22 were confidently identified. However, we have observed some limiting issues with regard to protein quantification using top-down MS/MS analyses of SILAC-labeled proteins. The role of SILAC labeling in the presence of competing endogenously produced amino acid residues and its impact on quantification of intact species are discussed in detail.

Introduction

Mass spectrometry based proteomics is a robust and rapidly developing field of interest seeking to understand “the identities, quantities, structures, and biochemical and cellular functions of all proteins in an organism, organ, or organelle, and how they vary in space, time, and physiological state”.[1] While the proteomics field is yet to achieve this ultimate

* **Author for Correspondence** W.M. Keck FT-ICR Mass Spectrometry Laboratory Department of Chemistry North Carolina State University Raleigh, North Carolina 27695 Phone: 919-513-0084 Fax: 919-513-7993 david_muddiman@ncsu.edu.

Supplementary Material Table S1 available for online access shows Identified proteins in addition to database accession information, intact protein masses, precursor ion masses, measured mass accuracy and P-score.

The data associated with this manuscript may be downloaded from Tranche (<http://tranche.proteomecommons.org>) using the following hash: PfBBXOHKrcjh9hm6pygUhqOIXvohLKq4dS8A/c5U3Qaj5HfAhG+M9YjN3hZVJJuNmKlwfK FgXKsV4sNEYvJjQ8cLP+4AAAAAAD+Q== The Tranche hash can also be used to verify that files have not changed since publication.

goal, two main strategies, termed “bottom-up” and “top-down”, are currently employed to pursue it. Bottom-up proteomics, a more widely implemented strategy relative to top-down, involves submitting a protein or protein mixture (e.g., cell lysate) to enzymatic digestion rendering the sample as a complex mixture of short peptides with defined C-termini. The peptides are then separated and analyzed via liquid chromatography (often reversed phase) coupled to electrospray mass spectrometry (ESI-MS) which sequences the peptides using tandem mass spectrometry (MS/MS). Searching the precursor ion mass and sequence information against a predicted protein database often results in hundreds or thousands of protein identifications.

However, sequence coverage of the original proteins is often low, with many identifications coming from the detection of a single peptide which may correspond to multiple protein forms.[2] These multiple protein forms may have unique biological function, potentially possessing diagnostic, prognostic, and/or therapeutic properties. Proteins such as brain natriuretic peptide, C-reactive protein, and troponin I and T are already regarded biomarkers in their native forms, but specific modifications (e.g. post-translational modifications (PTMs) and unique cleavages) can result in diagnostic markers that are more specifically related to a disease state.[3-8] Knowing the mass of the protein that gave rise to a specific proteolytic peptide would be very useful in determining its identity and whether it is a form related to a pathological stimulus.

Top-down proteomics provides both the intact mass and sequence information for protein identification and, though not employed as often as the bottom-up technique, continues to mature as a method for identifying and characterizing intact proteins in biological systems. [9] The intact protein mass is first obtained by the mass spectrometer, followed by isolation for MS/MS analysis to provide sequence information. The combination of intact mass and sequence data enable the specific form of the protein to be determined, locating PTMs and identifying alternative slicing variants and degradation products.

Top-down proteomics has lagged behind the more widely used bottom-up strategies due, in large part, to the technological challenges of measuring and identifying large intact proteins. These challenges have included time consuming off-line separation of samples[2, 10-12], mass analyzer speed[13], and underdevelopment of database search software tailored to top-down experiments. However, advances in separation technology and methodology[14, 15], the development of faster hybrid mass analyzer technology such as the LTQ-FT mass spectrometer[16], and recent innovations in top-down database searching software (e.g. ProSight PTM[17]) have created the potential for more widespread utilization of top-down proteomics techniques. Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI-FT-ICR-MS) is well suited for carrying out top-down proteomics measurements of intact multiply-charged proteins. The combination of high mass accuracy <3 ppm[18, 19], high resolving power, and a multitude of MS/MS dissociation techniques (e.g. CID[20], IRMPD[21], and ECD[22]) provide unprecedented confidence for protein identification and characterization (e.g. localization of PTMs, unique cleavages, and protein variants).[23, 24]

In addition to identifying and characterizing protein structures, understanding how they vary in quantity between different states is an integral goal of proteomic science. One of the first attempts at quantifying large, multiply-charged, intact proteins by Gordon et al. established that quantification could be achieved using relative abundances of an analyte and a structural analogue as an internal standard. However, a three amino acid difference in the amino acid sequence between equine and bovine cytochrome c resulted in hydrophobic differences between the analyte and internal standard. The resulting ESI bias produced two different

charge state distributions that limited quantification unless the abundances of all charge states were taken into account.[25]

Top-down analyses have also been used in conjunction with stable isotope labeling. Intact protein identification using amino acid counting and accurate mass has been accomplished using isotopic labeling of model cell cultures, including *E. coli* and *S. cerevisiae*, with deuterated leucine.[26, 27] However, deuterium labeling hinders quantification on a chromatographic scale as elution times vary between labeled (deuterium) and unlabeled analytes.[28, 29] Current intact protein quantification strategies currently employ stable isotopes that do not hinder liquid chromatography retention times. Kelleher and co-workers have used $^{15}\text{N}/^{14}\text{N}$ metabolic labeling to quantify the yeast proteome[30] and recently employed the technique using top-down LC-MS/MS to identify 22 proteins from *S. cerevisiae*. [31]

Ong and colleagues introduced stable isotope labeling of amino acids in cell culture (SILAC) as another means of quantifying peptides and proteins.[32, 33] Incorporation of ^{13}C and ^{15}N labeled amino acids using the SILAC technique results in co-eluting labeled and unlabeled species and has been widely implemented using isotopically labeled lysine and arginine in bottom-up experiments.[33, 34] Waanders and colleagues recently modeled and demonstrated top-down quantification and characterization of the signaling protein Grb2 expressed in *E. coli* using SILAC, with accurate quantification at 99% isotope enrichment. The protein was initially purified offline using Ni^{2+} affinity chromatography and analyzed using reversed phase liquid chromatography coupled to an LTQ-Orbitrap mass spectrometer. They also modeled the multiple isotopic envelopes that could result from incomplete incorporation of the heavy isotope amino acid with a 55 kDa protein.[35]

In this paper, we have assessed the performance of whole-organism SILAC labeling as a tool for top-down proteomics on the filamentous fungus *Aspergillus flavus* using a $^{13}\text{C}_6$ -arginine label and nLC-LTQ-FT-ICR-MS. *A. flavus* contaminates many plant species with the carcinogen aflatoxin resulting in economic losses in the hundreds of millions of dollars annually.[36] In addition to the agricultural impact of *A. flavus*, inhalation of conidia produced by the fungus can cause aspergillosis in patients with compromised immune systems. Secreted proteins of this organism has been previously examined by Medina et al., but were limited at that time by the absence of an annotated genome for the organism from which to identify many of their proteins.[37] A currently available annotated genome sequence for this fungus and the ability to easily culture this organism made it a good model for these studies.

SILAC labeling of this organism provides a powerful tool for measuring the biological effects of temperature, nutrition, and antifungal agents. As an example, this strategy will allow the identification, characterization, and quantification of regulatory proteins that are difficult of study by transcriptional profiling. Many regulatory proteins are constitutively expressed at a low level and are activated by protein localization or post-translational modification. Such regulatory elements are better studied by a proteomic analysis than by genomics.

This study represents our initial efforts to establish a top-down analysis of SILAC labeled *A. flavus*. We establish effective chromatography conditions for intact protein separation and discuss the limits and issues with amino acid incorporation in the culturing process.

Experimental

Culture of *A. flavus* for SILAC

Stock cultures of *Aspergillus flavus* strain NRRL 3357 were stored at -80°C in 35% glycerol. Inoculum was prepared by streaking an aliquot of the stock onto potato dextrose agar (PDA) (DIFCO; Lawrence, KS) supplemented with either i) $340\text{ }\mu\text{g/ml}$ $^{12}\text{C}_6$ -L-arginine (Sigma-Aldrich; St. Louis, MO) or ii) $340\text{ }\mu\text{g/ml}$ $^{13}\text{C}_6$ -L-arginine (Cambridge Isotopes; Andover, MA). After growth of the culture at 37°C for 5 days, conidia were dislodged from the cultures with a glass rod and suspended in 0.05% Triton X-100. The conidial suspension was quantified using a hemocytometer. Flasks containing 100 ml of A&M medium [38] supplemented with $340\text{ }\mu\text{g/ml}$ of either arginine $^{12}\text{C}_6$ or arginine $^{13}\text{C}_6$ were inoculated with conidial suspension to a final concentration of 10^6 conidia/ml. These liquid cultures were grown for 24 h at 28°C with shaking at 200 rpm allowing sufficient time for the conidia to germinate (around 6 to 7 hours) and grow for more than long enough to sufficiently incorporate the supplemented arginine (approximately 1 to 2 hours per cell cycle). Safety considerations: 1) Sample spores were treated as a potential respiratory hazard and handled in a LABGARD Laminar Flow Biological Safety Cabinet (Nuair, Inc Minneapolis, MN). 2) Samples were handled with gloves at all time to prevent and contamination from any aflatoxin present in the sample.

Protein Isolation and Sample Preparation

Fungal mycelium was separated from the medium by filtration through miracloth. The retained fungal mat was washed with cold PBS, removed from the filter, immediately frozen in liquid nitrogen, lyophilized, and stored at -80°C until extracted for proteins. A 0.2 g sample of lyophilized tissue was ground with a pestle in a mortar containing liquid nitrogen and 0.1 g of $150\text{--}200\text{ }\mu\text{m}$ glass beads (Sigma-Aldrich). The resultant fine powder was placed into 1.5 ml microcentrifuge tubes, resuspended with 1ml of cold PBS, and ground an additional 5 min at 4°C in a Vortex Genie (Scientific Industries, Bohemia, NY). The homogenate was spun for 5 min at $12,000 \times g$ to pellet large cellular debris. The supernatant was moved to a new tube and spun at $100,000 \times g$ for 45 min to pellet membranes and other debris. The resulting supernatant was submitted to acetone precipitation[39] to remove any residual detergents from culture. The pellet was resuspended in 20 mM monobasic potassium phosphate (pH 7.0) and passed through a 10 kDa MWCO filter. The resulting retentate (total protein) was quantified using a bicinchoninic acid (BCA) assay (BioRad, Hercules, CA) and utilized for SDS-PAGE and LC-MS experiments.

1D SDS-PAGE

Electrophoresis was performed using 12.5% Tris-HCl Criterion Gels (BioRad, Hercules, CA) in 25 mM Tris buffer containing 192 mM L-Glycine and 3.5 mM sodium dodecyl sulfate (SDS). $50\text{ }\mu\text{g}$ of the processed cell lysate ($> 10\text{ kDa}$) was loaded onto the gel and run at 150 V for 60 minutes and stained using 50 mL of Bio Safe Coomassie Blue G-250 (BioRad) according to the manufacturer's protocol.

Online LC-MS

All LC solvents were purchased from Burdick and Jackson (Muskegon, MI). Reversed phase liquid chromatography was performed using a $75\text{ }\mu\text{m}$ i.d. PicoFrit capillary column (New Objective, Woburn, MA) with a $15\text{ }\mu\text{m}$ emitter tip packed in-house with $5\text{ }\mu\text{m}$ mRP-C18 silica stationary phase (Agilent, Palo Alto, CA). The packed volume had dimensions $75\text{ }\mu\text{m}$ i.d. $\times 150\text{ mm}$ and was operated at room temperature. Samples ($1\text{ }\mu\text{g}$ total protein) were injected using a PAL Autosampler (LEAP Technologies, Carrboro, NC) and over the course of 10 minutes trapped and washed on a custom built mRP-C18 OPTI-PAK trap cartridge

(Optimize Technologies; Oregon City, OR) with 100% Mobile Phase A (95/5 water/acetonitrile) at 1 μ L/min until a 10 port switching valve (VICI, Houston, TX) was triggered to move the sample in-line with the gradient. Elution was carried out by a Chorus 220 nano-flow pump (CS Analytics, Zwingen, Switzerland) at 250 nL/min with mobile phases containing 95/5 (v/v) (Mobile Phase A) and 5/95 (Mobile Phase B) Water and Acetonitrile, respectively. The ion pairing reagent used was 0.2% formic acid (Sigma Aldrich, St. Louis, MO) in both mobile phases. The LC gradient was held at initial conditions of 10% B for 10 minutes followed by a ramp to 25% B over 5 minutes; 65% B was reached over the next 70 minutes, increased to 95% B in five minutes and held for an additional ten minutes before re-equilibrating at 5% B.

Mass spectrometry measurements were performed on a 7T LTQ-FT Ultra from Thermo Scientific (Bremen, Germany). The pulse sequence consisted of four events where mass measurements were all performed inside the ICR cell. The pulse sequence provided for a broadband acquisition in profile mode, followed by three data dependent MS/MS measurements. All events used 1 microscan to determine ionization time to reach the target AGC limit of 5×10^5 . The resolving power of all four events was set at 100,000_{FWHM} at $m/z=400$. MS/MS settings used an isolation width of 7 m/z and a normalized collision energy of 18% for a duration of 30 ms. CID was performed the three most abundant m/z values from the precursor ion scan in the LTQ followed by product ion detection in the ICR cell. Precursor ions with 1, 2, and 3 charges were excluded from MS/MS and dynamic exclusion for 60 seconds was used to reduce redundant analysis of the same precursors.

Data Analysis

RAW data files were processed manually, obtaining precursor m/z values from the MS/MS spectrum header and importing the MS/MS mass values into ProSightPC (Thermo Electron, Waltham MA) using a THRASH[40] algorithm to determine the monoisotopic peak of analyte signals with a S/N ratio of 5:1 or greater. The database was searched initially with absolute mass and biomarker search modes with precursor and fragment ion tolerances of ± 5 ppm and again with tolerances of ± 1 Da to accommodate sample handling modifications to the mass (e.g. deamidation) or error in monoisotopic peak picking by the THRASH algorithm. Sequences reported from the search were confirmed using arginine counting from the precursor ion spectra. Analysis of SILAC labeled protein distributions and theoretical distributions with varying amounts of arginine labeling was carried out using ICR-2LS (PNNL, Richland, WA) and Microsoft Excel® (Redmond, WA).

Results and Discussion

Two cultures of *Aspergillus flavus* were grown, one in standard medium and another in which $^{13}\text{C}_6$ -arginine was substituted for $^{12}\text{C}_6$ -arginine. After growth, mycelium from both cultures was lysed and the lysate combined in a 1:1 mixture based on total protein determination. After acetone precipitation the resuspended pellet was filtered through a molecular weight cut-off filter to remove detergents. After processing, the 1:1 lysate mixture of the two cultures was analyzed via nLC-LTQ-FT-ICR-MS. A representative base peak chromatogram showing total ion current of cell lysate samples processed from *A. flavus* is shown in Figure 1A. The left axis shows the relative total ion current and the right axis shows the % B of the mobile phase. The gradient profile represented by the dashed line spans 30% to 95% B and is superimposed upon the LC elution profile after being adjusted for dwell volume. Proteins producing precursor and product ion spectra eluted from 35 to 85 minutes, spanning most of the chromatographic gradient.

Figure 1B shows representative extracted ion chromatograms superimposed on the same chromatographic time scale axis. The separation efficiency of these species was evaluated

using the equation shown in the inset and serves as a metric to judge the performance of the separation and set a benchmark against which future separations can be compared.. The number of theoretical plates (N) was calculated using the exponentially modified Gaussian peak model developed by Foley and Dorsey[41], which takes account peak asymmetry (B/A) in addition to retention time (t_R) and peak width at 10% height ($W_{10\%}$). This model is well-suited for our investigations due to the asymmetric nature of the extracted chromatographic peaks. These peak shapes suggest non-ideal separation conditions that may include column temperature, ion pairing reagent used, and flow rate. Theoretical plate values ranged from 1.4×10^5 to 1.5×10^6 of the representative peaks having masses from 2.6 to 13.2 kDa with widths at 10% height ranging from 15 to 20 seconds.

The molecular weight distribution of the predicted proteome of *A. flavus* is shown in Figure 2 with a 0.1 kDa bin width. The *A. flavus* proteome database consisted of 12,847 predicted proteins [42]. The predicted proteins range in mass from 1.9 to 858 kDa, but for simplicity the 101 proteins exceeding 100 kDa are not shown. In a single LC-MS experiment, 1,318 unique species were observed spanning 1.9 to 18.7 kDa which are plotted in Figure 2 as a function of molecular weight. The same sample used in the LCMS experiment was also run on a one-dimensional SDS-polyacrylamide gel showing bands spanning the entire molecular weight range of the gel from 10 to 100 kDa. Product ion spectra from many of the low molecular weight precursor spectra were identified as protein fragments derived from proteins with significantly larger molecular weight when searched using ProSightPC. This suggested some *in vivo* processing or biological degradation of the sample from endogenous proteases. The overall low molecular weight distribution may also be attributed to electrospray bias for low molecular weight species when co-eluting with larger analytes, and/or chromatography biases.[14] The addition of other front end fractionation methods to further simplify the sample in addition to tailoring current instrument parameters including the ICR detection time, column temperature, gradient profile, and electrospray voltages may improve detection of larger intact species.

The top-down MS analysis of the SILAC labeled cell lysate mixture consisted of a precursor ion mass spectrum followed by three data dependent MS/MS measurements. Figure 3 shows three representative sets of MS and MS/MS spectra of unique SILAC protein pairs with their respective IDs using ProSightPC. The extracted ion chromatogram (XIC) of the monoisotopic masses of the SILAC pair is shown to the left in each example, demonstrating coelution of the pair. Figure 3A demonstrates the identification of a peptide sequence corresponding to a fragment of the glycolytic enzyme fructosebisphosphate aldolase in *A. flavus*. The isotopic envelopes believed to be a SILAC pair from the precursor ion spectrum were confirmed by examining their mass difference and extracted ion chromatograms of the heavy and light monoisotopic masses. The mass difference between the peaks corresponded to the incorporation of two $^{13}\text{C}_6$ -arginines into the heavy-labeled protein fragment. The lighter of the two peaks had an absolute abundance of 1×10^6 and was isolated for MS/MS resulting in a b- and y-ion series that was searched and identified using ProSightPC. Although the lighter peaks were not exclusively selected for collisional dissociation, confident identifications only arose from selection of the light peak as the search algorithm did not account for SILAC labeling and the mass difference of labeled arginines was too great to result in successful identification. Seventeen product ions from the MS/MS spectrum matched the predicted sequence giving an expectation value of 1.77×10^{-33} . In addition to the mass accuracy and the MS/MS sequence coverage, the identification of the protein was further confirmed using the number of predicted and observed arginines determined by mass difference of the SILAC pairs divided by the difference in heavy and light arginine (M/6.02013) in the precursor spectrum. Relative quantification of the heavy and light SILAC peak abundances corresponded strongly to a 1:1 ratio.

Figures 3B and 3C display protein fragments that were identified as HypA identical protein and Phosphoglycerate mutase family protein having absolute abundances of 2×10^6 and 3×10^6 , respectively. They were likewise confirmed as SILAC pairs using monoisotopic mass extracted ion chromatograms and identified using MS/MS spectra confirmed with arginine counting as previously described. Quantification was hindered, however, by the complexity of the heavy peak distributions which increased as the number of arginines in the protein increased. Despite the multiple isotopic distributions present in these complex heavy-labeled species, the light and heavy distributions were still regarded as a single SILAC pair. Incomplete incorporation of $^{13}\text{C}_6$ -arginine into the labeled *A. flavus* cultures due to the presence of endogenously produced arginine is the most likely cause resulting in these complex distributions. Overall, data-dependent MS/MS analysis was performed over two orders of magnitude of absolute ion abundance ranging from 1×10^4 to 3×10^6 and resulted in confident identification of 22 proteins (Supplemental Material Table S1).

The performance of SILAC on intact proteins depends greatly on the near-complete incorporation of the desired residue in the cultured organism. Figure 4 demonstrates the probability of complete incorporation of $^{13}\text{C}_6$ -arginine using Equation 1, where the abundance of heavy isotope labeled protein, A_H , depends on the $^{13}\text{C}_6$ -arginine labeling efficiency, E , raised to the exponent of the number of arginine residues to be labeled in the protein, N .

$$A_H = E^N \quad (\text{Equation 1})$$

The purity of supplemented $^{13}\text{C}_6$ -arginine (99.3%) is plotted along with the other efficiency values including the range observed in the top-down analysis of *A. flavus*. The model illustrates that the probability of complete SILAC labeling decreases with the number of arginines to be labeled in the protein (and proportionally as molecular weight increases). We have observed SILAC labeling efficiency of 72% to 85% in *A. flavus* suggesting incorporation of endogenously produced arginine during translation in cultures supplemented with $^{13}\text{C}_6$ -arginine. The chance of having a completely labeled 20 kDa protein at 72% $^{13}\text{C}_6$ -arginine labeling efficiency is 3.7%, which greatly hampers our ability to quantify any significant amount of the *A. flavus* proteome.

Incorporation of a heavy isotope label is modeled and compared to experimental broadband mass spectra, identified in Figure 3 using accurate intact mass, arginine counting, and MS/MS, by adjusting parameters in Equation 2, taking into account the total number of arginines in the protein, N , the labeling efficiency, E , and reporting relative abundances, A_n , for the masses of proteins containing n unlabeled arginines.

$$A_n = \frac{N!}{n!(N-n)!} E^n (1-E)^{N-n} \quad (\text{Equation 2})$$

Altering the labeling efficiency parameter influences the relative abundances of isotopic envelopes corresponding to differential $^{13}\text{C}_6$ -arginine incorporation. Representative spectra, shown in Figure 5A and B, which show in greater detail the precursor spectra from Figure 3A and 3B, identify closely with theoretical spectra calculated at 75% and 85% labeling efficiency, respectively, and is shown in the reflected axis below the experimental spectra. Circles overlaying the experimental spectra show the sum of model isotopic abundances from the different isotopic envelopes for more direct comparison between the model and the experimental data. This range of labeling efficiency agrees with labeling efficiency data obtained from tryptic digest experiments of the same sample performed to analyze relative protein expression at different growth temperatures using a bottom-up approach.[43] The disparity between the purity of heavy arginine supplemented for organism growth and the

amount actually incorporated into the proteome is an issue that can hinder both top-down and bottom-up quantification experiments. However, approaches such as reducing the amount of endogenous arginine incorporated during the culturing process or selecting an amino acid label not produced by the cultured organism could overcome this obstacle.

Conclusions

Using the SILAC technique, cultures of the filamentous fungus *Aspergillus flavus* were grown under identical conditions incorporating either $^{12}\text{C}_6$ or $^{13}\text{C}_6$ labeled arginine. Intact proteins from a 1:1 mixture of cell lysates from both cultures were then separated using an mRP stationary phase which produced 15 to 20 second peak widths in a nano-flow reversed phase HPLC coupled to an LTQ-FT-ICR mass spectrometer. In all, 1,318 intact or fragment protein masses were detected corresponding to 659 SILAC pairs, of which 22 were confidently identified. Broadband precursor ion spectra provided accurate intact mass and allowed for arginine counting which aided in the identification of proteins combined with the sequence information from data dependent MS/MS spectra. Quantification of proteins with few arginines was consistent with the 1:1 mixture ratio (Figure 3A). With greater numbers of arginine in a protein, increasingly complex heavy isotope spectra were observed, resulting in incorrect quantification. Endogenously produced arginine limited the labeling efficiency of the heavy culture to between 72% and 85% as determined using isotope distribution modeling. We are further developing methods to improve top-down quantification in SILAC experiments and expect to reduce endogenous amino acid incorporation through use of an arginine auxotroph in *A. flavus*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank LEAP Technologies for assistance with the nanoLC, Peter Mrozinski at Agilent for providing the mRP stationary phase, Professor Morteza Khaledi for insightful discussions regarding chromatography, and D.K. Williams Jr. for assistance with the 7T LTQ-FT. The authors gratefully acknowledge financial support received from the National Institutes of Health/NC State University Molecular Biotechnology Training Program (Grant 5T32GM008776-08), the W.M. Keck Foundation, and North Carolina State University.

References

- (1). Kenyon GL, DeMarini DM, Fuchs E, Galas DJ, Kirsch JF, Leyh TS, Moos WH, Petsko GA, Ringe D, Rubin GM, Sheahan LC. *Mol. Cell. Proteomics*. 2002; 1:763–780. [PubMed: 12438560]
- (2). Kelleher NL, Lin HY, Valaskovic GA, Aaserud DJ, Fridriksson EK, McLafferty FW. *J. Am. Chem. Soc.* 1999;806–812.
- (3). Hawkrigde AM, Heublein DM, Bergen RH III, Cataliotti A, Burnett JC Jr. Muddiman DC. *P. Natl. Acad. Sci. USA*. 2005; 102:17442–17447.
- (4). Das T, Sen A, Kempf T, Pramanik SR, Mandal C, Mandal C. *Biochem. J.* 2003; 373:345–355. [PubMed: 12693993]
- (5). Das T, Mandal C, Mandal C. *Glycoconjugate J.* 2004; 20:537–543.
- (6). Ardelt P, Dorka P, Jaquet K, Meilmeyer LMG, Kortke H, Korfer R, Notohamiprodjo G. *Biol. Chem.* 1998; 379:341–347. [PubMed: 9563831]
- (7). Katrukha AG, Bereznikova AV, Filatov VL, Esakova TV, Kolosova OV, Pettersson K, Lovgren T, Bulargina TV, Trifonov IR, Gratsiansky NA, Pulkki K, Voipo-Pulkki LM, Gusev N. *B. Clin. Chem.* 1998; 44:2433–2440.
- (8). Labugger R, Organ L, Collier C, Atar D, Van Eyk JE. *Circulation*. 2000; 102:1221–1226. [PubMed: 10982534]

- (9). Bogdanov B, Smith RD. *Mass Spectrom. Rev.* 2005; 24:168–200. [PubMed: 15389855]
- (10). Kelleher NL, Taylor SV, Grannis D, Kinsland C, Chiu HJ, Begley TP, McLafferty FW. *Protein Sci.* 1998; 7:1796–1801. [PubMed: 10082377]
- (11). Ge Y, Lawhorn BG, ElNaggar M, Strauss E, Park JH, Begley TP, McLafferty FW. *J. Am. Chem. Soc.* 2002; 124:672–678. [PubMed: 11804498]
- (12). Meng F, Cargile BJ, Patrie SM, Johnson JR, McLoughlin SM, Kelleher NL. *Anal. Chem.* 2002; 74:2923–2929. [PubMed: 12141648]
- (13). Siuti N, Kelleher NL. *Nat. Methods.* 2007; 4:817–821. [PubMed: 17901871]
- (14). Wang Y, Balgley BM, Rudnick PA, Lee CS. *J. Chromatogr. A.* 2005; 1073:35–41. [PubMed: 15909503]
- (15). Dillon TM, Bondarenko PV, Ricci MS. *J. Chromatogr. A.* 2004; 1053:299–305. [PubMed: 15543996]
- (16). Syka JEP, Marto JA, Bai DL, Horning S, Senko MW, Schwartz JC, Ueberheide B, Garcia B, Busby S, Muratore T, Shabanowitz J, Hunt DF. *Journal of Proteome Research.* 2004; 3:621–626. [PubMed: 15253445]
- (17). Zamborg LL, RD, Glowacz KJ, Kim Y, Viswanathan V, Spaulding IT, Early BP, Bluhm EJ, Babai S, Kelleher NL. *Nucleic Acids Res.* 2007; 35:W701–W706. [PubMed: 17586823]
- (18). Williams DK Jr, Muddiman DC. *Analytical chemistry.* 2007; 79:5058–5063. [PubMed: 17539605]
- (19). Nepomuceno AI, Mason CJ, Muddiman DC, Bergen HR, Zeldenrust SR. *Clin. Chem.* 2004; 50:1535–1543. [PubMed: 15205369]
- (20). Jennings KR. *Int. J. Mass Spectrom.* 2000; 200:479–493.
- (21). Little DP, Speir JP, Senko MW, O'Connor PB, McLafferty FW. *Anal. Chem.* 1994; 66:2809–2815. [PubMed: 7526742]
- (22). Zubarev RA, Kelleher NL, McLafferty FW. *J. Am. Chem. Soc.* 1998; 120:3265–3266.
- (23). Kelleher NL. *Anal. Chem.* 2004; 76:196A–203A.
- (24). Reid GE, McLuckey SA. *J. Mass Spectrom.* 2002; 37:663–675. [PubMed: 12124999]
- (25). Gordon EF, Mansoori BA, Carroll CF, Muddiman DC. *J. Mass Spectrom.* 1999; 34:1055–1062. [PubMed: 10510428]
- (26). Martinovic S, Veenstra TD, Anderson GA, Pasa-Toli L, Smith RD. *J. Mass Spectrom.* 2002; 37:99–107. [PubMed: 11813317]
- (27). Veenstra TD, Martinovic S, Anderson GA, Pasa-Toli L, Smith RD. *J. Am. Soc. Mass Spectrom.* 2000; 11:78–82. [PubMed: 10631667]
- (28). Zhang RJ, Sioma CS, Wang SH, Regnier FE. *Anal. Chem.* 2001; 73:5142–5149. [PubMed: 11721911]
- (29). Julka S, Regnier F. *J. Proteome Res.* 2004; 3:350–363. [PubMed: 15253416]
- (30). Du Y, Parks BA, Sohn S, Kwast KE, Kelleher NL. *Anal. Chem.* 2006; 78:686–694. [PubMed: 16448040]
- (31). Parks BA, Jiang L, Thomas PM, Wenger CD, Roth MJ, Boyne MT II, Burke PV, Kwast KE, Kelleher NL. *Anal. Chem.* 2007; 79:7984–7991. [PubMed: 17915963]
- (32). Ong S, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. *Mol. Cell. Proteomics.* 2002; 3:376–386. [PubMed: 12118079]
- (33). Ong S, Kratchmarova I, Mann M. *J. Proteome Res.* 2003; 2:173–181. [PubMed: 12716131]
- (34). de Godoy LMF, Olsen JV, de Souza GA, Li G, Mortensen P, Mann M. *Genome Biol.* 2006; 7:R50–R50.15. [PubMed: 16784548]
- (35). Waanders LF, Hanke S, Mann M. *J. Am. Soc. Mass Spectrom.* 2007; 18:2058–2064. [PubMed: 17920290]
- (36). Hussein HS, Brasel JM. *Toxicology.* 2001; 167:101–134. [PubMed: 11567776]
- (37). Medina MLH, PA, Brei L, Fancisco WA. *Proteomics.* 2005; 5:3153–3161. [PubMed: 16035112]
- (38). Mateles RI, Adye JC. *Appl. Microbiol.* 1965; 13:208–211. [PubMed: 14325880]
- (39). Jiang L, He L, Fountoulakis M. *J. Chromatogr. A.* 2004; 1023:317–320. [PubMed: 14753699]

- (40). Horn DM, Zubarev RA, McLafferty FW. J. Am. Soc. Mass Spectrom. 2000; 11:320–332. [PubMed: 10757168]
- (41). Foley JP, Dorsey JG. Anal. Chem. 1983; 55:730–737.
- (42). Payne GA, Nierman WC, Wortman JR, Pritchard BL, Brown D, Dean RA, Bhatnagar D, Cleveland TE, Machida M, Yu J. Med. Mycol. 2006; 44:S9–S11.
- (43). Georgianna DR, Hawkridge AM, Muddiman DC, Payne GA. J. Proteome Res. 2008 In Press.

Figure 1.

Base peak chromatogram (A) of reversed-phase separation of whole cell lysates from *A. flavus* and (B) Extracted ion chromatograms of peaks selected to analyze column performance using the Foley-Dorsey Equation (inset) which calculates theoretical plates, N , taking into account retention time, t_R , peak width at 10% height, $W_{10\%}$, and peak asymmetry, B/A .

Figure 2.

Histograms showing the molecular weight distributions in 0.1 kDa bins of the predicted proteome of *A. flavus* (top) and the observed molecular weights from a single LC-MS experiment (bottom). The distribution of molecular weight bands of intact species is visualized on a 1D-gel (right).

Figure 3.

Extracted ion chromatograms and MS spectra of SILAC pairs showing the corresponding elution of light and heavy isotopic envelopes of SILAC labeled proteins. Arginine counting and MS/MS spectra of the isolated light peak resulted in the identified sequence identified with confidence reported as an expectation value in ProSightPC.

Figure 4. Probability of complete heavy isotope incorporation into a given protein based on number of arginines at given labeling efficiencies. Using frequency of arginine in Swiss-Prot Database (~5%), number of Arginine residues are connected to a theoretical molecular weight.

Figure 5.

Full Scan FTMS spectra of SILAC pairs identified in Figure 3 and modeled heavy labeled isotope distributions using Equation 2 at labeling efficiencies best fitting the raw spectrum distribution. The sum of individual modeled isotopic peaks is overlaid upon the RAW spectra (circles).