Published in final edited form as:

J Am Soc Mass Spectrom. 2009 December ; 20(12): 2183–2191. doi:10.1016/j.jasms.2009.08.001.

A Robust Two-Dimensional Separation for Top-Down Tandem Mass Spectrometry of the Low-Mass Proteome

Ji Eun Lee^{a,*}, John F. Kellie^{a,*}, John C. Tran^a, Jeremiah D. Tipton^b, Adam D. Catherman^a, Haylee M. Thomas^a, Dorothy R. Ahlf^a, Kenneth R. Durbin^a, Adaikkalam Vellaichamy^a, Ioanna Ntai^a, Alan G. Marshall^{b,c}, and Neil L. Kelleher^a

^aUniversity of Illinois Urbana–Champaign, Department of Chemistry and The Institute for Genomic Biology, Urbana, Illinois, USA

^bIon Cyclotron Resonance Program, National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida, USA

^cDepartment of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida, USA

Abstract

For fractionation of intact proteins by molecular weight (MW), a sharply improved two-dimensional (2D) separation is presented to drive reproducible and robust fractionation before top-down mass spectrometry of complex mixtures. The “GELFrEE” (i.e., gel-eluted liquid fraction entrapment electrophoresis) approach is implemented by use of Tris-glycine and Tris-tricine gel systems applied to human cytosolic and nuclear extracts from HeLa S3 cells, to achieve a MW-based fractionation of proteins from 5 to >100 kDa in 1 h. For top-down tandem mass spectroscopy (MS/MS) of the low-mass proteome (5–25 kDa), between 5 and 8 gel-elution (GE) fractions are sampled by nanocapillary-LC-MS/MS with 12 or 14.5 tesla Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers. Single injections give about 40 detectable proteins, about half of which yield automated ProSight identifications. Reproducibility metrics of the system are presented, along with comparative analysis of protein targets in mitotic versus asynchronous cells. We forward this basic 2D approach to facilitate wider implementation of top-down mass spectrometry and a variety of other protein separation and/or characterization approaches.

Top-down mass spectrometry (MS) in which intact proteins are directly ionized and fragmented in the gas phase allows extensive characterization of the primary structure of a protein and provides the potential to characterize a variety of biological events that produce mass differences between mature proteins and the predicted products of their corresponding genes. Because top-down MS was initially used for characterizing single protein targets [1–3], steadily expanding efforts to extend the approach to complex proteome analysis have been hampered by the classic “front-end” problem of sample handling before MS. Effective fractionation of undigested proteome samples is critical to reduce sample complexity and obtain higher proteome coverage in top-down proteomics. Although two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) provides high peak capacity, no robust combination of 2D-PAGE with intact protein analysis by electrospray ionization (ESI)-MS has yet been reported. Thus,

© 2009 American Society for Mass Spectrometry. Published by Elsevier Inc.

Address reprint requests to Prof. Neil L. Kelleher, University of Illinois at Urbana-Champaign, Department of Chemistry, 53 Roger Adams Laboratory, 600 S. Mathews Avenue, Urbana, IL 61801, USA. kelleher@scs.uiuc.edu.

*These authors contributed equally to this work.

Appendix A Supplementary Material Supplementary material associated with this article may be found in the online version at doi: 10.1016/j.jasms.2009.08.001.

solution-phase separations have been the predominant option before top-down mass spectrometry. Proteome separations based on protein charge, such as ion-exchange, capillary isoelectric focusing (IEF), or chromatofocusing in conjunction with reversed-phase liquid chromatography (RPLC), have been demonstrated [4-7]. One-dimensional IEF or RPLC in capillaries has been coupled to ESI-Fourier transform (FT) MS for protein profiling [8-10]. Several years ago, our laboratory used gel-elution (GE) electrophoresis on a preparative scale [11,12] and we extend those studies here based partly on the work of Tran and Doucette [13].

In this study, we report an intact protein separation scheme based on molecular weight (MW) known as gel-eluted liquid fraction entrapment electrophoresis (GELFrEE) [13], referred to as GE here, as our first-dimensional protein fractionation. This molecular-weight-based separation involves continuous elution SDS-PAGE in a tube format, in which proteins are constantly eluted from the gel column and collected in the solution phase (i.e., free of the gel), providing broad mass range separation with high-resolution, reproducibility, and recovery. This GE technique generally provides rapid partitioning of a proteome into about 20 samples containing proteins in discrete mass ranges from <10 kDa to 200 kDa in 1 h. In our present study, 5–8 GE fractions containing up to 25-kDa proteins of human HeLa cell lysate were analyzed repeatedly by nanocapillary-RPLC online with 12 or 14.5 T LTQ-FT-MS for tandem mass spectrometry (MS/MS) on a chromatographic timescale. Using nuclear and cytosolic extracts from human cells in culture, we demonstrate all of the basic aspects of a prototypical workflow extendable to a proteomic scale, including comparative measurements of asynchronous and mitotic cells. Because a protein's mass predicts the fraction in which it will elute, this general approach should be applicable to a wide variety of problems in protein chemistry in both targeted and proteomic modes of operation.

Experimental

Preparation of Cytosolic and Nuclear Extracts

HeLa S3 cells obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA) were grown in suspension in Joklik's modified minimal essential medium supplemented with 10% newborn calf serum (NCS) and 100 U penicillin and streptomycin per mL. Upon reaching a density of 3×10^5 cells/mL, cells were then harvested by centrifugation at $200 \times g$ and washed twice with cold Tris-buffered saline (TBS). HeLa-S3 cell pellets consisting of 10^8 – 10^9 cells were resuspended with NIB-250 [15 mM Tris-HCl, pH 7.5, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 250 mM sucrose, 1 mM dithiothreitol, 10 mM sodium butyrate, 0.5 mM 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride, 50 nM microcystin plus 0.3% NP-40 at a 10:1 (vol/vol) ratio], and incubated on ice for 5 min. Nuclei and cytosol were separated by centrifugation at $600 \times g$ for 5 min and nuclei were twice rinsed with NIB-250 without NP-40. To generate histone-depleted nuclei, 500 μ L of NIB-250 plus 0.1% Rapigest (a gift from Waters, Milford, MA, USA) was added to the nuclear pellet and gently mixed. The mixture was incubated on ice for 5 min and then centrifuged at $14,000 \times g$ for 10 min at 4 °C. Protein concentrations were determined by use of the bicinchoninic acid (BCA, Rockford, IL, USA) method. Samples were reduced with 700 mM β -mercaptoethanol in SDS loading buffer and alkylated with iodoacetamide at an 80:1 M ratio at 25 °C in darkness for 60 min. M-phase arrested samples were prepared by adding 1 μ M colchicine to growing asynchronous cultures 3×10^5 cells/mL at 18 h before collection and were processed as described earlier.

GE Devices and Operating Conditions

A single-column GE device was fabricated similar to a device previously constructed [13]. For the reproducibility experiments, a multiplexed GE device was used [14]. The buffer systems were Tris-glycine (0.192 M glycine, 0.025 M Tris, 0.1% SDS) or Tris-tricine (0.1 M tricine, 0.1 M Tris, 0.1% SDS; Bio-Rad Laboratories, Inc., Hercules, CA, USA). Tube gels with Tris-

glycine and Tris-tricine media [15,16] were cast to 15% T for the resolving and 4% T for the stacking gels. Approximately 100 μ L of sample was loaded onto a GE column and 16 GE fractions with a volume of 150 μ L each were collected over 1 h after elution of the dye front. Operation of the device has been described previously [13].

Analytical SDS-PAGE Slab Gels

SDS-PAGE slab gel visualization of the collected GE fractions was used to assess separation. Ten μ L (6.7% of the total collected volume) of each fraction were loaded along with 5 μ L loading buffer onto a 15% T resolving slab gel. A standard protein ladder was loaded into one well of each slab gel. Gels were silver-stained and scanned with a traditional scanner [17].

Liquid Chromatography–Tandem Mass Spectrometry

GE fractions containing proteins from 5 to 25 kDa were cleaned to remove SDS as previously described [18]. Protein pellets were resuspended with 40 μ L of LC buffer A (95% H₂O, 5% acetonitrile, 0.2% formic acid), with protein recovery from the solubilization of the protein pellets at roughly 50%. Ten μ L of the sample was injected onto a 2-cm, 150- μ m i.d. C₄ or PLRP-S trap column (New Objective, Inc., Woburn, MA, USA). Samples were then washed onto a 10-cm, 75- μ m i.d. C₄ or PLRP-S analytical column (New Objective) with either a short or long gradient flowing at 300 nL/min from an Eksigent 1D Plus system (Eksigent, Dublin, CA, USA). The short gradient started at 95% buffer A and 5% buffer B (5% H₂O, 95% acetonitrile, 0.2% formic acid) and continued to 20% B in 10 min, 60% B in 45 min, and 95% B in 10 min. The long gradient started at 5% buffer B and continued to 30% B in 10 min, 40% B in 60 min, 60% B in 20 min, and 95% B in 15 min.

Samples were analyzed with either a commercial 12 T or modified 14.5 T LTQ-FT-Ultra mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA). Both instruments were calibrated with a standard Thermo calibration mixture consisting of caffeine, the peptide MRFA, and Ultramark. Parameters for the 12 T LTQ-FT-Ultra were a data-dependent top 2 MS/MS acquisition strategy with m/z 10 isolation window, 8 microscans at 180 K resolving power ($m/\Delta m_{50\%}$, in which $\Delta m_{50\%}$ is mass spectral peak full width at half-maximum peak height) in the MS scan event (m/z 500–2000 scan range), and 8 microscans at 90 K resolving power in the MS/MS scan events (scan ranges from one third of the precursor m/z to m/z 2000). Dynamic exclusion was enabled with a repeat count of 1, an exclusion duration of 5000 s, and a repeat duration of 240 s. Secondary ion accumulation with the 14.5 T LTQ-FT-MS has been described previously [19]. Parameters for the 14.5 T LTQ-FT-MS were as follows: data-dependent top 3 MS/MS acquisition strategy with m/z 14 isolation window, 4 cycles from the LTQ to the wired octopole [20], 2 microscans at 200 K resolving power in the MS scan event (m/z 500–2000 scan range), 10 cycles of secondary ion accumulation, followed by a single scan at 200 K resolving power in the MS/MS scan event (scan ranges from one third of the precursor m/z to m/z 1800). Dynamic exclusion was enabled with a repeat count of 2, an exclusion duration of 400 s, and a repeat duration of 60 s.

Data Analysis

Each GE-LC-MS/MS file was analyzed by use of ProSightPC 2.0 (Thermo Fisher Scientific). First, intact precursor and fragment masses were determined and compiled into a human-readable .xml file called a .puf file (ProSight Upload Format). The .puf file was then searched against a shotgun annotated [21] human proteome database containing known post-translational modifications (PTMs), coding single nucleotide polymorphisms (cSNPs), and alternative splice forms via an iterative search tree method. The first absolute mass search used a 100-Da precursor window and 10 parts per million (ppm) fragment mass tolerances. For initial searches that did not identify a protein below an E-value cutoff of 10^{-4} , a second absolute mass search took place with a 10,000-Da precursor window. The two search modes were used

for computational efficiency of searching on ProSightPC 2.0. The 100-Da window allowed the expeditious identification of proteins with modifications in a small mass range such as cSNPs and unknown PTMs, whereas the 10,000-Da window identified proteins with modifications in a large mass range, including multiple unknown PTMs and alternative splicing [22]. The second part of the search tree also used “Delta M mode,” a feature that applies the mass difference between the experimental and candidate mass to all the fragment ions such that they match a mass shift not known a priori within 10 ppm. Protein forms with borderline E-values ($\sim 10^{-3}$ or higher) were either searched further or examined manually to improve identification confidence.

Custom Software and Computer Programming

Software for intact mass determination from an LC-MS file has been described previously [23]. Files were processed with Online Automation cRAWler to determine masses detected from GE-LC-MS injections and intact masses from technical triplicate injections. Heat maps were generated by a Microsoft Windows XP application written in the C# programming language by use of the Microsoft .NET 2.0 Framework and developed in the Microsoft Visual Studio 2005 environment. The MLApp.dll was used to integrate MATLAB 7.0 (The MathWorks, Natick, MA, USA) into the application. A new computer program, Proteome Display, was devised for visualization of nano-LC-MS/MS data, allowing detailed viewing of individual protein forms in individual fractions. The protein forms from each fraction were retrieved from the corresponding Automation Warehouse (SQL database populated by Online Automation cRAWler) [23]. The m/z of the protein form was used to extract its selected ion chromatogram (SIC). An SIC shows the temporal change in abundance of ions for a specified mass range during chromatography, with each GE fraction represented by an x/y plot of retention time versus intact mass and a heat map with color scaled according to ion relative abundance.

Results and Discussion

GE Buffer Systems

The GE device is capable of high-resolution separation of intact proteins, with a typical GE run displayed in the inset of the overall workflow shown in Figure 1 (top right). Here, we see a Tris-glycine separation of the 5- to 25-kDa portion of the proteome into roughly five fractions. Tris-tricine was also used and achieved slightly higher resolution in the 5- to 40-kDa size regime, even with a 2-fold greater protein load (see Supplemental Figure S1, which can be found in the electronic version of this article). Comparing analogous fractions from Tris-glycine and Tris-tricine GE runs revealed that, on average, the number of protein masses detected shrunk from about 50 to 40. Below roughly 25 kDa, Tris-tricine produces about eight useful fractions, whereas Trisglycine produces five or six (Supplemental Figure S1). As a result, Tris-tricine provides samples with a greater focus on proteins in the <25-kDa mass range versus that of Tris-glycine, but requires more nanocapillary RPLC injections to analyze this portion of the proteome. There are differences in the two buffer systems as MW increases (as can be seen in the figures); however, the detection and identification metrics at high MW for each buffer system have not been extensively analyzed. Other SDS-PAGE buffer systems, such as Tris-acetate, are suitable for work at high mass. As the upper mass limit of reliable top-down proteomics by use of GE-LC-MS/MS identification increases, such buffer systems will become practical.

LC-MS/MS Performance

Favorable metrics for sample throughput and data analysis are vital to the practicality of top-down proteomics. Although the nature of our automated data analysis is well established [21], critical assessment of data acquisition in high-throughput mode is required. Each GE

fraction <25 kDa was subjected to nano-LC-MS/MS. Each of the 4–8 GE fractions contained about 10 μ g total protein per fraction before sample cleanup. Figure 2 shows representative data and ProSight output from an LC-MS/MS injection of fraction #3 of a Tris-glycine GE run for HeLa nuclear extract with data collection at 14.5 T. Figure 2a shows a base-peak chromatogram for this GE fraction, with representative MS and MS/MS spectra (with identifications and E-values) illustrated in Figure 2b and c, respectively. Highlighted in Figure 2d is an N-terminal acetylated form of nucleoside diphosphate kinase B, identified with an E-value of 10^{-42} . In total, 37 distinct protein forms were detected in this run. There were 71 redundant (total) protein identifications corresponding to 13 unique accession numbers. These data were typical for this study and highlight one challenge of top-down mass spectrometry: the same protein form was repeatedly targeted for fragmentation within a single injection. Thus, a simple but significant improvement would be to prevent the instrument from fragmenting different charge states of the same protein, a need not fulfilled by traditional m/z -based dynamic exclusion.

In Figure 2a, chromatographic peak widths (full width at half-maximum, or FWHM) fell in the 15- to 45-s range. Such chromatography is favorable from a separations perspective, but places a temporal constraint on MS/MS data acquisition. Although LTQ-based peptide analysis can afford >60% MS/MS spectra leading to no identification [24], each MS/MS scan recorded at FT-MS resolution event is “more important,” given that relatively long data acquisition periods for FT-MS/MS result in roughly 3- to 5-fold fewer MS and MS/MS datasets per unit time with current instrumentation. For top-down MS, high mass accuracy in MS/MS is highly advantageous for precise characterization of PTMs, but is acquired at the expense of duty cycle. The cycle time achieved during the top 2 or 3 data-dependent MS/MS acquisitions is roughly 4–5 cycles/min (~2–6 s/scan). For that cycle time, each abundant protein is targeted for fragmentation at least once. Within any given elution peak, whole protein charge-state distributions are observed (Figure 2b) routinely with isotopic resolution.

Another representative GE-LC-MS/MS injection for 12 T LTQ-FT-MS of GE fraction #5 can be seen in Supplemental Figure S2, with 50 protein forms detected and 25 of these identified and listed in Table 1. Further, the metrics for data acquisition and protein identification by use of ProSight were collected (Supplemental Table S1). Approximately one quarter of the MS/MS events led to an identification, with the proteins ranging from 5 to 12.5 kDa and identified with a median E-value of 10^{-16} .

Considering seven distinct GE-LC-MS/MS injections from one GE run, 286 distinct protein forms were detected, resulting in 225 redundant (total) protein identifications, corresponding to identification of 35 unique protein accession numbers. The number of identified protein forms per GE-LC-MS/MS injection varied from 10 to 20 unique protein identifications (i.e., unique accession numbers). This straightforward implementation of the present GE-LC-MS/MS platform enabled a significant increase in the average number of protein identifications compared to the platforms previously reported by our laboratory [11,12].

LC-MS Visualization

One way to evaluate protein masses detected with the GE-LC-MS/MS platform is to simply plot the detected masses as a function of LC elution time for each GE fraction, which provides a visual metric of GE-LC-MS platform performance. This type of graphical comparison between slab gel visualization of solution-phase GE fractions to masses observed in the mass spectrometer is shown in Figure 3. In-house software for intact mass determination was used for the MW range from 3 to about 35 kDa, corresponding to the first 9 GE fractions; in all, 453 distinct protein masses were detected by ESI-FT-MS at 12 T across these 9 fractions (data not shown). In general, the intensity of regions on the silver-stained gel correlates well to mass values detected by MS, as one would expect. While serving as a validation of nano-LC-MS/

MS and sample handling after GE, the main uses of LC-MS data visualization lie in assessing both sample coverage for a single run and absolute changes in signal magnitude between multiple GE runs for different cell states (see following text), providing sufficient analytical precision.

GE-LC-MS Reproducibility

The reproducibility of a high-throughput proteomic platform is critical for uncovering protein level dynamics in complex systems. Beyond the standard uses for technical or biological replicates in proteomics, top-down proteomics projects in their early phases will require repeat analyses in targeted mode for two main reasons: to identify protein forms putatively related to those already found or to fully characterize proteins that are identified but need their modifications precisely localized. In either context, reproducibility has special value in initial application of top-down proteomics to a new proteome.

To assess reproducibility, analyses of GE runs and nano-LC-MS were performed in technical triplicates. For GE alone, reproducibility across six GE separations was assessed via slab gels (see Supplemental Figure S3). The six GE lanes were run simultaneously via a multiplexed GE device [14]. Most of the same protein bands can be seen in precisely the same fraction on all six slab gels. A key element in reproducible GE operation is the casting of the stacking and resolving gels. The GE system recently became commercially available, so pre-cast gels made in bulk should give higher reproducibility than the self-cast gels used here. Excellent chromatographic reproducibility for three injections of a 10- to 15-kDa GE fraction was also observed (Supplemental Figure S4a). Highly reproducible charge-state distributions at the same retention time for each of the three injections were also observed (Supplemental Figure S4b). The reproducibility of charge state abundance can be important for data-dependent MS/MS, although it is perhaps expected given the same protein concentrations and ESI conditions during each run. By use of our in-house intact mass determination algorithm, a total of 54 masses overlapped in all three technical triplicate injections (Supplemental Figure S4c). This is encouraging for top-down proteomic analyses; however, further reproducibility studies will be necessary to fully analyze the reproducibility of the GE-LC-MS/MS platform when implemented on a proteomic scale.

Differences in Asynchronous and M-Phase Arrested HeLa Cells

Initial implementation of this newly developed GE-LC-MS/MS platform focused on the comparison of low-mass proteins from asynchronous and M-phase arrested HeLa cells. For example, various histone forms have been previously characterized by use of top-down mass spectrometry, with quantification of their modification stoichiometries [24-27]. Here, we highlight specific examples of phosphorylation level changes between asynchronous cells and those that were mitotically arrested. Figure 4 displays two proteins, stathmin and 60S acidic ribosomal protein, for which phosphorylation peaks (+80 Da) differentially increased or decreased between the two cell states. Although the phosphorylation levels of stathmin largely increased in M-phase arrest cells, those of 60S acidic ribosomal protein decreased during mitosis. Stathmin is an important regulatory protein of mitotic microtubule dynamics [28] and the phosphorylation of stathmin is known to occur during the G2/M transition [29]. Four phosphorylations on serine residues of stathmin (Ser16, Ser25, Ser38, and Ser63) are known to increase during M-phase progression; they promote dissociation of stathmin from microtubules [28,29]. Given that this protein is abundant, oncogenic, and has been the subject of biochemical studies for over two decades, the hierarchy (i.e., order) of phospho-Ser site usage is putatively known; this system deserves targeted testing by top-down MS/MS but is beyond the scope of the current study. The context for phosphorylation of the 60S acidic ribosomal protein (ARP) is also known [30-31]. Although ARP is known to participate in translational regulation by interacting with eukaryotic translation elongation factor 2,

phosphorylation of ARP can affect ribosomal activity by increasing the affinity for the ribosome. Our current approach allows us to measure how the relative abundances of non-, singly-, and doubly-phosphorylated forms of the protein change in asynchronous and mitotic cells. Direct measurement of modification stoichiometries in this fashion complements bottom-up MS and antibody-based measurements to provide deeper insight of understanding PTM dynamics.

Conclusion

For fractionation of intact proteins, the described GE approach is robust and general. We expect a diversity of separation approaches to be implemented with little difficulty before implementation of the GE/nano-LC system described here. Further, a variety of postseparation measurements can be made beyond or besides top-down MS, including bottom-up MS/MS, Western blotting, or manufacturing of protein arrays. For our laboratory's immediate interest, the extension of the methods outlined here to higher mass proteins is a part of a multifaceted approach to lower the main barrier to top-down proteomics: proteome coverage. As the GELC-MS/MS platform is proven extensible to other labs and streamlined further, we believe that other instruments such as Q-TOFs [32-34], LTQ-Orbitraps [35], or even stand-alone ion traps [36-38] could be used for low MW proteome analysis based on our GE separation. We also hope that others will adopt top-down or hybrid bottom-up/top-down approaches, both of which can benefit substantially from MW-based separations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Christopher L. Hendrickson, Mark R. Emmett, John P. Quinn, and Greg T. Blakney of the National High Magnetic Field Laboratory for helpful advice and discussion, and the National Science Foundation (DMR-06-54118) for supporting this national facility. The authors also thank Mingxi Li for helping with sample preparations. The laboratory of NLK was supported by the Packard Foundation, the Sloan Foundation, and the National Institutes of Health (GM-067193-07) and JL was supported by the UIUC Neuroproteomics Center on Cell-Cell Signaling supported by the National Institute on Drug Abuse (P30 DA-018310-06).

References

1. Kelleher NL, Costello CA, Begley TP, McLafferty FW. Thiaminase-I (42-kDa) Heterogeneity, Sequence Refinement, and Active-Site Location from High-Resolution Tandem Mass-Spectrometry. *J. Am. Soc. Mass Spectrom* 1995;6:981–984.
2. Kelleher NL, Lin HY, Valaskovic GA, Aaserud DJ, Fridriksson EK, McLafferty FW. Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc* 1999;121:806–812.
3. McLafferty FW. High-Resolution Tandem FT Mass-Spectrometry above 10-kDa. *Acc. Chem. Res* 1994;27:379–386.
4. Chong BE, Yan F, Lubman DM, Miller FR. Chromatofocusing Nonporous Reversed-Phase High-Performance Liquid Chromatography/Electrospray Ionization Time-of-Flight Mass Spectrometry of Proteins from Human Breast Cancer Whole Cell Lysates: A Novel Two-Dimensional Liquid Chromatography/Mass Spectrometry Method. *Rapid Commun. Mass Spectrom* 2001;15:291–296. [PubMed: 11223961]
5. Lubman DM, Kachman MT, Wang HX, Gong SY, Yan F, Hamler RL, O'Neil KA, Zhu K, Buchanan NS, Barder TJ. Two-Dimensional Liquid Separations—Mass Mapping of Proteins from Human Cancer Cell Lysates. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci* 2002;782:183–196.

6. Sharma S, Simpson DC, Tolic N, Jaitly N, Mayampurath AM, Smith RD, Pasa-Tolic L. Proteomic Profiling of Intact Proteins Using WAX-RPLC 2-D Separations and FTICR Mass Spectrometry. *J. Proteome Res* 2007;6:602–610. [PubMed: 17269717]
7. Zhou F, Johnston MV. Protein Profiling by Capillary Isoelectric Focusing, Reversed-Phase Liquid Chromatography, and Mass Spectrometry. *Electrophoresis* 2005;26:1383–1388. [PubMed: 15759303]
8. Jensen PK, Pasa-Tolic L, Peden KK, Martinovic S, Lipton MS, Anderson GA, Tolic N, Wong KK, Smith RD. Mass Spectrometric Detection for Capillary Isoelectric Focusing Separations of Complex Protein Mixtures. *Electrophoresis* 2000;21:1372–1380. [PubMed: 10826683]
9. Lee SW, Berger SJ, Martinovic S, Pasa-Tolic L, Anderson GA, Shen YF, Zhao R, Smith RD. Direct Mass Spectrometric Analysis of Intact Proteins of the Yeast Large Ribosomal Subunit Using Capillary LC/FTICR. *Proc. Natl. Acad. Sci. U.S.A* 2002;99:5942–5947. [PubMed: 11983894]
10. Yang LY, Lee CS, Hofstadler SA, Pasa-Tolic L, Smith RD. Capillary Isoelectric Focusing-Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Protein Characterization. *Anal. Chem* 1998;70:3235–3241. [PubMed: 11013724]
11. Du Y, Meng FY, Patrie SM, Miller LM, Kelleher NL. Improved Molecular Weight-Based Processing of Intact Proteins for Interrogation by Quadrupole-Enhanced FT MS/MS. *J. Proteome Res* 2004;3:801–806. [PubMed: 15359734]
12. Meng FY, Cargile BJ, Patrie SM, Johnson JR, McLoughlin SM, Kelleher NL. Processing Complex Mixtures of Intact Proteins for Direct Analysis by Mass Spectrometry. *Anal. Chem* 2002;74:2923–2929. [PubMed: 12141648]
13. Tran JC, Doucette AA. Gel-Eluted Liquid Fraction Entrapment Electrophoresis: An Electrophoretic Method for Broad Molecular Weight Range Proteome Separation. *Anal. Chem* 2008;80:1568–1573. [PubMed: 18229945]
14. Tran JC, Doucette AA. Multiplexed Size Separation of Intact Proteins in Solution Phase for Mass Spectrometry. *Anal. Chem* 2009;81:6201–6209.
15. Schagger H. Tricine-SDS-PAGE. *Nat. Protoc* 2006;1:16–22. [PubMed: 17406207]
16. Schagger H, Vonjagow G. Tricine Sodium Dodecyl-Sulfate Polyacrylamide-Gel Electrophoresis for the Separation of Proteins in the Range from 1-Kda to 100-Kda. *Anal. Biochem* 1987;166:368–379. [PubMed: 2449095]
17. Shevchenko A, Wilm M, Vorm O, Mann M. Mass Spectrometric Sequencing of Proteins from Silver Stained Polyacrylamide Gels. *Anal. Chem* 1996;68:850–858. [PubMed: 8779443]
18. 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;138:141–143. [PubMed: 6731838]
19. Schaub TM, Hendrickson CL, Horning S, Quinn JP, Senko MW, Marshall AG. High-Performance Mass Spectrometry: Fourier Transform Ion Cyclotron Resonance at 14.5 Tesla. *Anal. Chem* 2008;80:3985–3990. [PubMed: 18465882]
20. Wilcox BE, Hendrickson CL, Marshall AG. Improved Ion Extraction from a Linear Octopole Ion Trap: SIMION Analysis and Experimental Demonstration. *J. Am. Soc. Mass Spectrom* 2002;13:1304–1312. [PubMed: 12443021]
21. Zamdborg L, LeDuc RD, Glowacz KJ, Kim YB, Viswanathan V, Spaulding IT, Early BP, Bluhm EJ, Babai S, Kelleher NL. ProSight PTM 2.0: Improved Protein Identification and Characterization for Top Down Mass Spectrometry. *Nucleic Acids Res* 2007;35:W701–W706. [PubMed: 17586823]
22. Roth MJ, Forbes AJ, Boyne MT, Kim YB, Robinson DE, Kelleher NL. Precise and Parallel Characterization of Coding Polymorphisms, Alternative Splicing, and Modifications in Human Proteins by Mass Spectrometry. *Mol. Cell. Proteomics* 2005;4:1002–1008. [PubMed: 15863400]
23. Wenger CD, Boyne MT, Ferguson JT, Robinson DE, Kelleher NL. Versatile Online-Offline Engine for Automated Acquisition of High-Resolution Tandem Mass Spectra. *Anal. Chem* 2008;80:8055–8063. [PubMed: 18841935]
24. Elias JE, Haas W, Faherty BK, Gygi SP. Comparative Evaluation of Mass Spectrometry Platforms Used in Large-Scale Proteomics Investigations. *Nat. Methods* 2005;2:667–675. [PubMed: 16118637]
25. Pesavento JJ, Bullock CR, Leduc RD, Mizzen CA, Kelleher NL. Combinatorial Modification of Human Histone H4 Quantitated by Two-Dimensional Liquid Chromatography Coupled with Top Down Mass Spectrometry. *J. Biol. Chem* 2008;283:14927–14937. [PubMed: 18381279]

26. Pesavento JJ, Mizzen CA, Kelleher NL. Quantitative Analysis of Modified Proteins and Their Positional Isomers by Tandem Mass Spectrometry: Human Histone H4. *Anal. Chem* 2006;78:4271–4280. [PubMed: 16808433]
27. Roth MJ, Parks BA, Ferguson JT, Boyne MT, Kelleher NL. “Proteotyping”: Population Proteomics of Human Leukocytes Using Top Down Mass Spectrometry. *Anal. Chem* 2008;80:2857–2866. [PubMed: 18351787]
28. Belmont LD, Mitchison TJ. Identification of a Protein That Interacts with Tubulin Dimers and Increases the Catastrophe Rate of Microtubules. *Cell* 1996;84:623–631. [PubMed: 8598048]
29. Larsson N, Melander H, Marklund U, Osterman O, Gullberg M. G2/M Transition Requires Multisite Phosphorylation of Oncoprotein-18 by 2 Distinct Protein-Kinase Systems. *J. Biol. Chem* 1995;270:14175–14183. [PubMed: 7775478]
30. Naranda T, Ballesta JPG. Phosphorylation Controls Binding of Acidic Proteins to the Ribosome. *Proc. Natl. Acad. Sci. U.S.A* 1991;88:10563–10567. [PubMed: 1961721]
31. Vidales FJ, Robles MTS, Ballesta JPG. Acidic Proteins of the Large Ribosomal-Subunit in *Saccharomyces cerevisiae*: Effect of Phosphorylation. *Biochemistry* 1984;23:390–396. [PubMed: 6421316]
32. Deterding LJ, Lhattacharjee S, Ramirez DC, Mason RP, Tomer KB. Top-Down and Bottom-Up Mass Spectrometric Characterization of Human Myoglobin-Centered Free Radicals Induced by Oxidative Damage. *Anal. Chem* 2007;79:6236–6248. [PubMed: 17637042]
33. Nemeth-Cawley JF, Tangarone BS, Rouse JC. “Top Down” Characterization is a Complementary Technique to Peptide Sequencing for Identifying Protein Species in Complex Mixtures. *J. Proteome Res* 2003;2:495–505. [PubMed: 14582646]
34. Whitelegge JP, Zhang HM, Aguilera R, Taylor RM, Cramer WA. Full Subunit Coverage Liquid Chromatography Electrospray Ionization Mass Spectrometry (LCMS+) of an Oligomeric Membrane Protein: Cytochrome b6f Complex from Spinach and the Cyanobacterium *Mastigocladus laminosus*. *Mol. Cell. Proteomics* 2002;1:816–827. [PubMed: 12438564]
35. Macek B, Waanders LF, Olsen JV, Mann M. Top-Down Protein Sequencing and MS3 on a Hybrid Linear Quadrupole Ion Trap-Orbitrap Mass Spectrometer. *Mol. Cell. Proteomics* 2006;5:949–958. [PubMed: 16478717]
36. Bowers JJ, Liu J, Gunawardena HP, McLuckey SA. Protein Identification via Ion-Trap Collision-Induced Dissociation and Examination of Low-Mass Product Ions. *J. Mass Spectrom* 2008;43:23–34. [PubMed: 17613176]
37. Bunger MK, Cargile BJ, Ngunjiri A, Bundy JL, Stephenson JL. Automated Proteomics of *E. coli* via Top-Down Electron-Transfer Dissociation Mass Spectrometry. *Anal. Chem* 2008;80:1459–1467. [PubMed: 18229893]
38. Chi A, Bai DL, Geer LY, Shabanowitz J, Hunt DF. Analysis of Intact Proteins on a Chromatographic Time Scale by Electron Transfer Dissociation Tandem Mass Spectrometry. *Int. J. Mass Spectrom* 2007;259:197–203. [PubMed: 17364019]

Figure 1.

Workflow for top-down proteomics. Total protein content from HeLa-S3 nuclei or cytosol is quantified and loaded onto a gel-elution (GE) column. The GE device separates the protein samples according to molecular weight (MW). Proteins of increasing MW elute into solution-phase fractions, which can be visualized on a slab gel (top right). The solution-phase fractions are cleaned up to remove SDS before injection onto a nano-liquid chromatography (LC) column for tandem mass spectrometric (MS/MS) analysis with an LTQ-FT at either 12 or 14.5 T. LC-MS/MS files are processed with ProSightPC 2.0, a software suite tailored for top-down analysis in a high-throughput setting.

Figure 2.

Examples selected from an LC-MS/MS injection of fraction #3 from a Tris-glycine GE run. (a) A base-peak chromatogram is shown, with (b) charge state distributions from selected retention times. (c) Abundant charge states (above the arrows) were targeted for fragmentation. Fragmentation mass spectra for each protein are shown along with the corresponding identifications and E-values. A fragmentation map (d) results from the matching fragment ions of nucleoside diphosphate kinase B. The protein is N-terminally acetylated.

Figure 3.

Analytical slab gel and heat map showing masses detected in GE fractions. The silver-stained slab gel (top) of solution-phase fractions shows typical GE separation (Tris-tricine buffer system). After LC-MS of the GE fractions, the detected masses for each fraction are plotted as a function of LC retention time (bottom). The MS relative abundance is designated with the scale on the right; the inset illustrates the enlarged region of heat map of GE fraction 6. General correlation can be seen between detected masses and abundances from the MS and the intensities of protein bands detected on the slab gel.

Figure 4.

Differences in phosphorylation for proteins identified in asynchronous and M-phase arrested HeLa cells. (a) Asynchronous HeLa proteins are identified with the phosphorylations designated in red. (b) M-phase arrested HeLa proteins are identified, again with phosphorylations designated in red. Dynamic changes in phosphorylation between asynchronous and M-phase arrested HeLa cells were clearly observed. (c) A fragment map for the doubly phosphorylated 60S acidic ribosomal protein (phosphorylation sites in red text).

Table 1

Proteins identified from a single GE-LC-MS/MS injection ^a

Protein description	Observed mass (Da)	Calculated mass (Da)	Mass difference (ppm)	PTMs	Matching fragments	E-value	$-10 \times \log$ (E-value)
Migration-inducing protein 12	4933.5	4933.5	0.2	N-terminal acetylation	32	1×10^{-36}	360
40S ribosomal protein S30	6643.8	6643.8	0.1		27	4×10^{-34}	334
Metallothionein-2	6932.6	7046.6	N/A ^b		16	5×10^{-19}	183
Small EDRK-rich factor 2	6933.6	6937.7	N/A ^b	N-terminal acetylation	9	1×10^{-6}	60
Coiled-coil domain-containing protein 72	6972.9	6972.9	-1.1	N-terminal acetylation	13	3×10^{-21}	205
40S ribosomal protein S28	7935.2	7935.2	0.3	N-terminal acetylation	29	5×10^{-46}	453
60S ribosomal protein L38	8138.7	8138.7	-4.5		8	4×10^{-8}	74
Oncogene DJ-1 (fragment)	8264.3	8264.3	-0.6		8	3×10^{-10}	95
Ubiquitin (fragment)	8845.6	8845.6	0.7		61	6×10^{-96}	952
GAPDH (fragment)	8855.6	8855.6	0.5		13	9×10^{-16}	159
1-AGP acyltransferase 2 (fragment)	9252.9	9252.8	11.2		9	1×10^{-6}	60
ATPase inhibitory factor 1 (fragment)	9510.9	9510.9	0.5		16	2×10^{-9}	87
S100 calcium-binding protein A6	10,141.3	10,141.3	1.4	N-terminal acetylation	10	3×10^{-10}	95
U6 snRNA-associated SM-like protein LSM8	10,307.3	10,307.3	0.8	N-terminal acetylation	44	2×10^{-79}	787
Dynein-associated protein Km23	10,825.7	10,825.7	1.1	N-terminal acetylation	10	3×10^{-13}	125
Heat shock 10-kDa protein	10,835.7	10,835.8	-5.7	N-terminal acetylation	79	3×10^{-104}	1035
S100 calcium-binding protein A13	11,375.1	11,375.1	0.6	N-terminal acetylation	22	2×10^{-28}	277
S100 calcium-binding protein A11	11,757.8	11,757.8	1.3	N-terminal acetylation	29	2×10^{-17}	167
S100 calcium-binding protein A16	11,761.0	11,762.0	12.3	N-terminal acetylation	21	2×10^{-32}	317
SUMO-1-specific protease 2 (fragment)	11,785.7	11,785.2	32.3		5	6×10^{-6}	52
S100 calcium-binding protein A4	11,794.7	11,859.8	N/A ^b	N-terminal acetylation	9	5×10^{-6}	53
Thioredoxin	11,883.8	11,883.8	-0.5	Methylation	11	1×10^{-15}	150
Glutaredoxin-1	11,964.3	11,964.3	1.0	Methylation	15	6×10^{-14}	132

Protein description	Observed mass (Da)	Calculated mass (Da)	Mass difference (ppm)	PTMs	Matching fragments	E-value	$-10 \times \log$ (E-value)
Phenylpyruvate tautomerase	12,508.2	12,508.2	1.3		25	3×10^{-34}	335
MIF protein	12,538.3	12,551.2	N/A ^b		8	1×10^{-5}	50

^aThe base peak chromatogram for this run is shown in Supplemental Figure S2.

^bThe observed protein forms were not present in the database; therefore, large mass discrepancies exist between the observed and calculated forms.