Published in final edited form as:

Proteomics. 2011 July ; 11(14): 2931–2935. doi:10.1002/pmic.201100045.

Comparison between Procedures using Sodium Dodecyl Sulfate for Shotgun Proteomic Analyses of Complex Samples

Michael S. Bereman, Jarrett D. Egertson, and Michael J. MacCoss

Department of Genome Sciences, University of Washington, Seattle, WA

Abstract

Filter aided sample preparation (FASP) and a new sample preparation method using a modified commercial SDS removal spin column are quantitatively compared in terms of their performance for shotgun proteomic experiments in three complex proteomic samples: a *Saccharomyces cerevisiae* lysate (insoluble fraction), a *Caenorhabditis elegans* lysate (soluble fraction), and a human embryonic kidney cell line (HEK293T). The characteristics and total number of peptides and proteins identified are compared between the two procedures. The SDS spin column procedure affords a conservative 4-fold improvement in throughput, is more reproducible, less expensive (i.e., requires less materials), and identifies between 30–107% more peptides at a q 0.01, than the FASP procedure. The peptides identified by SDS spin column are more hydrophobic than species identified by the FASP procedure as indicated by the distribution of GRAVY scores. Ultimately, these improvements correlate to as great as a 50% increase in protein identifications with 2 or more peptides.

Keywords

Bottom-up proteomics; shotgun proteomics; protein identifications; sample preparation protocols; sodium dodecyl sulfate

Shotgun proteomics is a powerful strategy to survey the protein content of complex mixtures. This field has progressed significantly in the last decade due to advancements in MS instrumentation[1–4], bioinformatics[5–7], as well as the continued refinement of sample preparation methodologies. The major steps of a shotgun proteomic experiment include: 1) sample preparation; 2) peptide separation (e.g., HPLC); 3) mass detection; and 4) data analysis. The success of an experiment is predicated on following a robust, reproducible sample preparation method. A robust sample preparation method lends statistical power to comparative experiments (i.e., biomarker studies) by being reproducible (i.e., lower replicate variance) and high throughput (i.e., greater number of samples analyzed per unit time).

Although laboratories will have a preferred method for sample preparation of protein mixtures for LC-MS analysis, typically all procedures will follow these common steps including: 1) solubilization and denaturation of proteins; 2) reduction of disulfide bonds; 3) alkylation of reduced cysteines; 4) digestion; and 5) sample cleanup. The identification of any single protein in a complex mixture is highly dependent on the efficacy of the digestion step which is correlated directly to the ability to dissociate protein complexes, denature tertiary structures, and to solubilize hydrophobic membrane-bound proteins. Urea, a

* **Author for Correspondence:** Michael J. MacCoss, Ph.D., Department of Genome Sciences, University of Washington, Seattle, WA, Phone: 206.616.7451, maccoss@uw.edu.

The authors have declared no conflict of interest.

chaotrope, is commonly used for protein denaturation by disrupting non-covalent bonds between atoms. Although urea is readily removed prior to MS by RP-HPLC, it does have distinct disadvantages.[8] Other methods for protein denaturation include the use of recently developed acid-cleavable detergents which can be removed prior to mass spectrometry by simply decreasing the pH of the sample – a necessary step regardless. Although these detergents are effective and MS-compatible[8–9], the most well characterized and widely used detergent by far in the biological sciences is sodium dodecyl sulfate (SDS). A major pitfall in the use of ionic detergents (e.g., SDS) is the detrimental effects of these species on LC-MS experiments making it necessary to remove them prior to analysis. Removal of SDS has been performed using organic solvent precipitations[10]; however, these techniques can be characterized as low throughput, time consuming, and high sample loss – especially of hydrophobic proteins that are known to partition into organic solvents[11]. Due to the continued and growing use of SDS in proteomic related experiments (i.e., gel eluted liquid fraction entrapment[12] GELFrEE), it is essential that robust methods to remove SDS from proteomics mixtures be explored.

Herein, we evaluate a new method for shotgun proteomics in which proteins are digested in the presence of 0.1% SDS and then SDS is removed prior to LC-MS using a modified commercial detergent removal spin column. The amount of stationary phase in the spin column is optimized for high sample recovery and reproducibility from simple and a complex mixture of proteins. We then compare the method to a recently presented procedure to remove high concentrations of SDS – termed filter aided sample preparation (FASP)[13] which is based off an earlier method reported by Manza *et al.*[14] We demonstrate that the simple SDS spin column method offers a considerable increase in throughput and protein identifications across a diverse set of complex proteomic mixtures.

Filter-aided sample preparation (FASP) was performed using a 10 kDa MWCO filter as reported previously.[13] This procedure was compared to a method using a modified commercial SDS removal spin column to deplete the sample of SDS after digestion. Proteins were solubilized in 50 mM ammonium bicarbonate and 0.1% SDS. Reduction, alkylation, and digestion were performed in the presence of SDS. Trypsin is reported to retain a majority of its activity in the presence of small amounts of SDS.[15] After digestion and prior to LC-MS/MS analysis, SDS was removed from the samples using a self packed spin column. The procedure used in these studies for packing the spin columns is described in detail in supplemental information.

Three complex proteomic samples were used to evaluate the sample preparation procedures described *vide supra* including a *S. cerevisiae* lysate (insoluble fraction), *C. elegans* lysate (soluble fraction), and a human embryonic kidney cell line (HEK293T). Procedures for preparation of each sample type can be found in supplemental information. Approximately 200 µg of total protein for each sample, determined using a commercial bicinchoninic acid (BCA) assay (Thermo Fisher Scientific), were prepared in triplicate using both sample preparation procedures. Common to samples prepared by both procedures were the use of heat (95° C) for denaturation and a 4 hour trypsin digestion. For the human embryonic kidney cell line all samples were subjected to short periods of sonication (5 × 30 s) to aide lysis.

Nano-flow liquid chromatography was performed using an Agilent 1100 quaternary pump configured in a vented column design[16]. ESI columns were prepared as reported previously[6] and then packed to a length of 50 cm with 4 µm C12 reverse phase particles (Phenomenex, Torrance, CA). A 4 hour total LC-MS analysis was performed using a gradient with a linear ramp from 9% B to 38% B over 185 minutes, the gradient increased to 80% B over 5 minutes and held constant for 10 minutes, and then the column was re-

equilibrated for 20 mins at initial conditions (91% A/9% B). Approximately 2 μ g of protein material were injected on column. Mass spectral analyses were performed using a 2-dimensional linear ion trap (Thermo Fisher Scientific, San Jose, CA) equipped with an electrodynamic funnel.[17] ESI-MS parameters are described in detail elsewhere.[16] Peptide matches were assigned to spectra using an in-house developed pipeline and the Sequest algorithm. Experimental spectra were searched against a sequence (target) and reversed-sequence (decoy) database of all annotated proteins for the appropriate organism. Candidate peptides for each spectrum were restricted to a ± 3 Da precursor mass tolerance, and included peptides resulting from semi-tryptic digestion and/or a single missed cleavage event. The target and decoy database search results were processed by Percolator[5] to improve peptide-spectrum matches and enforce a peptide level q value threshold of 0.01. Gene Ontology (GO) annotations were assigned using ProteinCenterTM from Proxeon. Further description of the analysis pipeline can be found in the supplemental material.

The two sample preparation procedures explored in these studies are summarized in Figure 1. The FASP protocol is relatively time-consuming with 10 major steps and a significant amount of time dedicated to sample handling/processing. The other protocol, referred to as the SDS spin column procedure, minimizes sample handling and makes use of a modified commercial spin column to remove SDS from the sample post-digestion. Ultimately this procedure results in a conservative 4-fold increase in throughput when compared to FASP and is less expensive because it requires less materials. Furthermore, our SDS spin column procedure actually performs the digestion in SDS as opposed to just solubilization in SDS and digestion in urea. Herin, we quantitatively compare these procedures in terms of reproducibility, total number and characteristics of peptides/proteins identified across a diverse set of complex proteomic samples.

Figure 2A–C displays the number of peptides identified as a function of q value amongst the different sample types for the three technical replicates of the respective procedures. The SDS spin column procedure yielded a greater number of peptides identified at any particular q value (e.g., 0–0.1) for all samples investigated. The average number of peptides identified (q 0.01) from the three technical replicates is compared in Figure 2D–F. On average the SDS spin column procedure identified statistically a greater number of peptides for each sample type at the 95% confidence level ($p < 0.05$) using a 1-tailed t test. In addition, the number of peptide spectral matches (q 0.01) using the SDS spin procedure were significantly greater (supplemental Figure 2).

Figure 2G–I displays a histogram comparing the Kyte and Doolittle grand average hydropathy score[18] (i.e., a measure of hydrophobicity) for the peptides identified (q 0.01) between the procedures for each sample type. All histograms are biased slightly towards more negative values (i.e., hydrophilic species) which is in agreement with previous shotgun proteomic studies[19]. Although the FASP procedure utilizes a large amount of detergent for solubilizing sample, surprisingly this did not correlate to a higher number of hydrophobic peptides identified. On average the SDS spin column procedure identified more hydrophobic peptides – indicated by the slight shift of the histogram towards zero. The biggest discrepancy in hydrophobicity is observed in the human cell line, which itself is interesting because it is likely that a higher concentration of detergent would afford more efficient lysis and solubilization of hydrophobic membrane-bound proteins. These observations may be explained by either the loss of hydrophobic peptides due to the numerous sample preparation steps in FASP or the hydrophobic proteins potentially could precipitate in the MWCO cartridge following SDS removal prior to enzymatic digestion.

The last metric used to evaluate the peptides identified between the two procedures is referred to as the propensity to buried (PTB)[20]. This metric identifies residues that are typically located on the inside (i.e., buried) or outside of a protein's tertiary structure by examining each amino acid's likelihood to have contact with water. The greater the PTB score the more amino acid residues the peptide possesses that typically reside on the inside of the protein. Figure 2J–L display histograms of the PTB scores for the different organisms between the two procedures. Histograms were created by combining all peptides identified with $q < 0.01$ from the technical replicates. Although, very similar distributions are shown for both procedures, the SDS spin column procedure on average identified peptides that have slightly larger PTB scores than the peptides identified from FASP – indicating these peptides consist of amino acids that are more likely to reside inside a protein structure. Similar to the GRAVY scores, the biggest difference in this metric between the two sample preparation methods was observed in the human cell line data.

Figure 3 displays the average number of proteins and indistinguishable groups identified amongst the sample types using 1 and 2 peptides. An indistinguishable group is defined by proteins that are identified by the same peptide(s), and thus, it is impossible to determine the presence of any single or group of proteins unambiguously.[21] It is important to examine these numbers as simply reporting the total number of proteins identified can be misleading. For example, over 5000 proteins are identified using the spin column procedure in the kidney cell line; however, there are only ~2200 protein groups. The SDS spin column procedure identifies on average between 5–20% and 10–50% more proteins than the FASP for the various sample types using 1 and 2 peptide identifications, respectively. The SDS spin procedure is more reproducible in the number of protein identifications given the smaller standard deviations (i.e., error bars) in the mean of the technical replicates. We attribute this higher reproducibility to the simplicity and limited number of preparation steps associated with the SDS spin column procedure.

Finally, we investigated the possibility that a bias may exist in the type (e.g., cellular component) of proteins identified between the two procedures by comparing the GO annotations. Neither procedure differed significantly in the percentage of protein identified from various cellular components; however, on average the SDS spin column procedure identified a greater number of membrane proteins (Supplemental Figure 3). This discrepancy in membrane proteins was most evident in the *C. elegans* and human cell data.

In this study, a shotgun proteomics procedure using 0.1% SDS for in-solution digestion and a modified spin column for SDS removal has been systematically evaluated and compared to the FASP procedure. The SDS spin column, as presented here, yields a considerable improvement in throughput along with increased number of peptide and protein identifications from three different complex proteomic mixtures. This method will be of considerable use for future bottom-up experiments where SDS is required in the sample preparation[12] and throughput is essential to handle moderate to large numbers of sample.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors greatly acknowledge the financial support from the National Institutes of Health (R01 DK069386, U01 HG004263, and P41 RR011823). MSB acknowledges support from the Genome Training Grant (T32 HG000035). The authors are greatly appreciative for the *S. cerevisiae* and *C. elegans* samples provided by Jennifer Merrihew. Veronika Glukhova is thanked for the HEK293T cells. Vagisha Sharma is acknowledge for aid with the GO Annotations.

References

1. Syka JEP, Marto JA, Bai DL, Horning S, et al. Novel linear quadrupole ion trap/FT mass spectrometer: Performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res.* 2004; 3:621–626. [PubMed: 15253445]
2. Hu QZ, Noll RJ, Li HY, Makarov A, et al. The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* 2005; 40:430–443. [PubMed: 15838939]
3. Hardman M, Makarov AA. Interfacing the orbitrap mass analyzer to an electrospray ion source. *Anal Chem.* 2003; 75:1699–1705. [PubMed: 12705605]
4. Second TP, Blethrow JD, Schwartz JC, Merrihew GE, et al. Dual- Pressure Linear Ion Trap Mass Spectrometer Improving the Analysis of Complex Protein Mixtures. *Anal Chem.* 2009; 81:7757–7765. [PubMed: 19689114]
5. Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semisupervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods.* 2007; 4:923–925. [PubMed: 17952086]
6. Hsieh EJ, Hoopmann MR, MacLean B, MacCoss MJ. Comparison of Database Search Strategies for High Precursor Mass Accuracy MS/MS Data. *J Proteome Res.* 2010; 9:1138–1143. [PubMed: 19938873]
7. Mayampurath AM, Jaitly N, Purvine SO, Monroe ME, et al. DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. *Bioinformatics.* 2008; 24:1021–1023. [PubMed: 18304935]
8. Chen EI, Cociorva D, Norris JL, Yates JR. Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J Proteome Res.* 2007; 6:2529–2538. [PubMed: 17530876]
9. Norris JL, Merrihew G, Yi X, MacCoss MJ. Improving proteome analysis using cleavable surfactants. *Molecular & Cellular Proteomics.* 2006; 5:748.
10. Botelho D, Wall MJ, Vieira DB, Fitzsimmons S, et al. Top-Down and Bottom-Up Proteomics of SDS-Containing Solutions Following Mass-Based Separation. *J. Proteome Res.* 2010
11. Beers MF, Bates SR, Fisher AB. Differential Extraction for the Rapid Purification of Bovine Surfactant Protein-B. *Am J Physiol.* 1992; 262:L773–L778. [PubMed: 1616060]
12. 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]
13. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nature Methods.* 2009; 6:359–U360. [PubMed: 19377485]
14. Manza LL, Stamer SL, Ham AJL, Codreanu SG, Liebler DC. Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics.* 2005; 5:1742–1745. [PubMed: 15761957]
15. Porter WH, Preston JL. Retention of trypsin and chymotrypsin proteolytic activity in sodium dodecyl sulfate solutions. *Anal Biochem.* 1975; 66:69–77. [PubMed: 1147220]
16. Klammer AA, MacCoss MJ. Effects of modified digestion schemes on the identification of proteins from complex mixtures. *J. Proteome Res.* 2006; 5:695–700. [PubMed: 16512685]
17. Page JS, Tang K, Smith RD. An electrodynamic ion funnel interface for greater sensitivity and higher throughput with linear ion trap mass spectrometers. *Int J Mass Spectrom.* 2007; 265:244–250.
18. Kyte J, Doolittle RF. A Simple Method for Displaying the Hydropathic Character of a Protein. *J Mol Biol.* 1982; 157:105–132. [PubMed: 7108955]
19. Chen EI, Cociorva D, Norris JL, Yates JR. Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J Proteome Res.* 2007; 6:2529–2538. [PubMed: 17530876]
20. Wertz DH, Scheraga HA. Influence of Water on Protein-Structure - Analysis of Preferences of Amino-Acid Residues for inside or Outside and for Specific Conformations in a Protein Molecule. *Macromolecules.* 1978; 11:9–15. [PubMed: 621952]
21. Zhang B, Chambers MC, Tabb DL. Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *J. Proteome Res.* 2007; 6:3549–3557. [PubMed: 17676885]

Figure 1.

A comparison of filter aided sample preparation (FASP) and SDS spin column procedures for shotgun proteomics.

Figure 2.

Data describing peptides identified from *S. cerevisiae* (1st column), *C. elegans* (2nd column) and human (3rd column) for FASP (blue) and SDS spin column procedures (red). Peptides vs. q value plots for the different organisms are listed in A–C. All three technical replicates are displayed. The average number of peptides from the replicates identified with a $q = 0.01$ is listed for the respective organisms in D–F. Error bars correspond to 95% CI. Histograms for the GRAVY scores and propensity to be buried are listed in G–I and J–L, respectively. The histograms were created by combining all peptides identified with a $q = 0.01$ from the individual replicates for each organism.

Figure 3.

The number of proteins identified between the two procedures for **A)** *S. cerevisiae*; **B)** *C. elegans*; and **C)** human. Proteins identified matching to at least 1 peptide, 2 peptides and their respective indistinguishable groups are listed. Error bars correspond to the standard deviation in the mean.