Published in final edited form as:
J Proteome Res. 2007 July ; 6(7): 2529–2538. doi:10.1021/pr060682a.

Optimization of Mass Spectrometry Compatible Surfactants for Shotgun Proteomics

Emily I. Chen[1], **Daniel Cociorva**[1], **Jeremy L. Norris**[2], and **John R. Yates III**[1,*]

1*Department of Cell Biology, 10550 North Torrey Pines Road, SR11, The Scripps Research Institute, La Jolla, CA 92037*

2*Protein Discovery, Inc. Knoxville, TN 37902*

Abstract

An optimization and comparison of trypsin digestion strategies for peptide/protein identifications by μLC-MS/MS with or without MS compatible detergents in mixed organic-aqueous and aqueous systems was carried out in this study. We determine that adding MS compatible detergents to proteolytic digestion protocols dramatically increases peptide and protein identifications in complex protein mixtures by shotgun proteomics. Protein solubilization and proteolytic efficiency are increased by including MS compatible detergents in trypsin digestion buffers. A modified trypsin digestion protocol incorporating the MS compatible detergents consistently identifies over 300 proteins from 5ug of pancreatic cell lysates and generates a greater number of peptide identifications than trypsin digestion with urea when using LC/MS/MS. Furthermore, over 700 proteins were identified by merging protein identifications from trypsin digestion with three different MS compatible detergents. We also observe that the use of mixed aqueous and organic solvent systems can influence protein identifications in combinations with different MS compatible detergents. Peptide mixtures generated from different MS compatible detergents and buffer combinations show a significant difference in hydrophobicity. Our results show that protein digestion schemes incorporating MS compatible detergents generate quantitative as well as qualitative changes in observed peptide identifications, which lead to increased protein identifications overall and potentially increased identification of low abundant proteins.

Keywords

shotgun proteomics; mass spectrometry; proteolysis; cancer cells; protein profiling

Introduction

Mass spectrometry has been a key technology of proteomic based biological discovery. Two approaches to protein analysis involve either “bottom up” – pre-digestion of intact proteins prior to mass spectrometry analysis or “top down” – direct mass spectrometry analysis of intact proteins [1, 2]. When “bottom up” is applied to protein mixtures it is often known as “shotgun proteomics”. Four important steps are involved in shotgun proteomics: 1) protein digestion, 2) peptide separation, 3) peptide fragmentation in the mass spectrometer, and 4) database searching (data analysis)[3-5]. The initial step of protein digestion is often the most critical step as it involves reduction of intact proteins into a collection of peptides of suitable size for mass spectrometry analysis. Proteins exist within cells as tightly associated collections of proteins

*Correspondence: John R. III Yates, jyates@scripps.edu, Phone: (858) 784-8862, Fax (858) 784-8883, The Scripps Research Institute, 10550 North Torrey Pines Road, SR 11, La Jolla, CA92037.

(complexes) or inserted into lipid bilayers [6]. Effective digestion of proteins often requires dissociation of complexes and denaturation of proteins as well solubilization of hydrophobic, lipid-bilayer bound proteins. Urea is a chaotrope used to disrupt intraprotein interactions for more efficient digestion of proteins [7]. Urea is often used to assist in proteolytic digestion of proteins since it is readily removed by reversed-phase liquid chromatography. Several disadvantages exist in the use of urea for protein denaturation. Proteases are subject to the denaturation effects of urea so a concentration of urea that still retains protease activity must be identified and this concentration is often below the desirable concentration to fully denature the protein mixture [8]. Under the influence of heat, urea can decompose to isocyanic acid and carbamylate free amines in peptides and proteins. Carbamylation can block sites of trypsin digestion or create artefactual modifications of peptides that are difficult to distinguish from naturally occurring modifications such as acetylation and trimethylation. An alternative to urea is the use of detergents like sodium dodecyl sulfate (SDS) to denature and solubilize proteins, but detergents are notoriously incompatible with reversed-phase liquid chromatography and mass spectrometry and must be removed prior to analysis.

To improve denaturation of proteins for mass spectrometry-based analyses, mass spectrometry-compatible detergents have been developed. An acid-labile surfactant (ALS), sodium 3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxyl]-1-propanesulfonate (I) or RapiGest SF, was first described as an SDS analog to facilitate direct analysis of intact proteins by mass spectrometry [2]. To create mass spectrometry compatibility, ALS is hydrolyzed at low pH to create molecules that minimize interference with reversed-phase separations and ionization [9]. Sine then, ALS has been utilized to improve in-gel [10] and in-solution protein digestions [11] prior to mass spectrometry and liquid chromatography analysis. Studies using ALS have demonstrated that unlike SDS, ALS solubilizes proteins and shows no significant inhibition of trypsin activity up to 1% (w/v) [11]. Furthermore, ALS can be used to solubilize hydrophobic proteins and improve in-solution digestion of proteolytically resistant proteins [11, 12]. Finally, the acid-labile nature of the surfactant allows rapid and simple sample cleanup prior to LC-MS or LC-MS/MS analysis of complex protein mixture.

In our study, we compare the utility of three commercially available mass spectrometry compatible detergents: RapiGest SF (RPG), Invitrosol (IVS), and PPS Silent Surfactant (PPS). RapiGest SF and PPS Silent Surfactant are acid-labile surfactants. Invitrosol (IVS) is a homogeneous surfactant that elutes in three peaks well separated from the elution times of most peptides. Similar to previous studies with ALS, we observe that all three surfactants can act as mild denaturants to increase protein solubilization in solution. Also, they can be used to solubilize and unfold hydrophobic proteins at concentrations that do not denature or inhibit common proteolytic enzymes such as trypsin. In addition to comparing the utility of these MS compatible surfactants for in-solution protein digestion, we also examine the effect of organic and aqueous solvents on protein digestion in the presence of MS compatible surfactants. Mixtures of organic solvents and water have been shown to influence the digestion of proteins [13]. Many proteases are active in reasonably high concentrations of organic solvents such as methanol and acetonitrile. In some cases the amino acid residue specificity is altered or proteolysis rates are increased in the presence of high concentrations of organic solvents. High concentrations of organic solvents can also disrupt protein folding and denature proteins. In this paper we study the effectiveness of mass spectrometry (MS) compatible detergents in combination with high concentrations of organic solvents to promote the digestion of complex mixtures of proteins.

Results and Discussions

MS compatible detergents improve protein solubility and proteolytic efficiency of trypsin in solution

Shotgun proteomic experiments using LC/MS/MS utilize in-solution digestion of protein mixtures. Proteolytic digestion of complex protein mixtures such as mammalian cell lysates and tumor tissues, however, is often limited by protein solubility in solution and the folded states of proteins. Protein complexes present in protein mixtures often require detergents to denature complexes to make the proteins more susceptible to proteolysis. The amphiphilic nature of detergents encourages unfolding and dissolution of hydrophobic proteins. However, most commonly used detergents in biological applications such as SDS and Triton X-100 have to be removed prior to downstream mass spectrometry analysis because detergents frequently interfere with binding, elution, and ionization of peptides during LC-MS/MS separation. To eliminate detergent interference, several MS compatible detergents have been developed for applications in mass spectrometry analysis. They are acid-cleavable detergents, RapiGest SF (RPG) and PPS, as well as the non-cleavable detergent, Invitrosol. Hence, we set out to optimize and evaluate the benefits of incorporating MS compatible detergents to enzymatic digestion schemes for complex protein mixtures.

Critical functions of detergents are solubilization and denaturation of proteins. Both are important for proteolysis of proteins. We examined the ability of the three detergents to improve solubility relative to 2 M urea. Catalase is a hydrophobic protein that is insoluble in aqueous solution. 5μg aliquots of Catalase were solubilized either in an MS compatible detergent (RPG, PPS, and INV), or 2 M urea in the Tris buffer. The insoluble Catalase was centrifuged and the solubilized Catalase was then analyzed using SDS-PAGE with coomassie staining. The density of the stained bands was measured and we observed that all three MS compatible detergents as well as 2M urea increased solubility of Catalase by a factor of 1.8 to 5.0 relative to buffer without detergents or urea (Figure 1A). Additionally, proteolytic efficiency in the presence of an MS compatible detergent was measured. To assess proteolytic efficiency independent from protein solubility, a soluble protein, bovine serum albumin (BSA) was used as a substrate and digestion by trypsin was measured over a period of 5 hours by determining the amount of undigested protein. Ten micrograms of BSA was incubated with or without each detergent and incubated at 37°C for five hours. 10μl aliquots were removed every hour and analyzed by SDS-PAGE followed by coomassie staining. By visualizing the intensity of full length BSA from the stained gel the amount of undigested BSA could be determined. After 5hr incubation with trypsin in the aqueous buffer with no detergent a significant amount of full length BSA remained undigested compared to control BSA without trypsin. However, over 90 percent of the full length BSA disappeared after 4 hours in BSA digested with trypsin and MS compatible detergents (Figure 1B). We conclude the MS compatible detergents improve digestion of BSA.

Trypsin activity is enhanced in mixed organic–aqueous systems

Russell et al. reported the activity of trypsin is increased in the presence of organic-aqueous solvents for the digestion of single proteins [13]. They did not test the use of trypsin in the presence of organic solvents and other chaotropes for complex protein mixtures, but presumably the organic solvent served as a denaturant during digestion. While organic solvents can denature proteins they do not have a strong solubilzing effect and denatured proteins can precipitate from solution rendering the increased activity of trypsin ineffective. The activity of trypsin toward digestion of 5ug of pancreatic cancer cell lysates was measured in mixed organic-aqueous solvent systems consisting of 80% acetonitrile, 80% 2-propanol, and 80% methanol, respectively. The measure of trypsin activity in the digestion of complex protein mixtures was based on the number of peptides identified using LC/MS/MS. Peptide identifications obtained by trypsin digestion in the mixed organic-aqueous solvents were

compared with results of trypsin digest in aqueous buffer (100mM Tris-HCl, pH 8.0). We found the number of peptides identified decreased from a high of 967 (322 proteins) to a low of 181 (157 proteins) in the presence of 80% acetonitrile, 80% 2-propanol, 100mM Tris, and 80% methanol, respectively (Table 1). The results show that digestion efficiency of complex proteins mixtures is improved over aqueous solutions as indicated by the number of peptides identified using LC/MS/MS.

MS Compatible Detergents Increase Peptide and Protein Identifications in the Mixed Organic-Aqueous and Aqueous Systems

Five micrograms of soluble protein extracted from pancreatic cancer cells was digested with trypsin in aqueous 2M and 8M urea buffer. A solution of RPG (0.1%), IVS (1X) and PPS (0.2%) was mixed with organic-aqueous (80% acetonitrile) and aqueous buffers (Tris-HCl, pH 8.0) for digestion of the protein mixture. Digestion was quantified by the number of peptides identified and subsequently the number of proteins. Only invitrosol (IVS) was used for subsequent comparisons in other mixed organic-aqueous systems. All analyses were carried out in triplicate with the same amount of protein and analyzed with an identical experimental setup. Tandem mass spectra collected in each experiment were searched using SEQUEST against a mouse IPI protein database plus a decoy database (see Method section). Results from all digestion conditions were evaluated based on five parameters: total number of peptides identified, tryptic status, total number of proteins identified, number of unique protein identifications across all conditions, and physical properties of peptides identified.

A. MS compatible detergents increase the total number of identified peptides— We observe a significant improvement in the number of peptides and proteins identified with the incorporation of MS compatible detergents to the digestion buffer. Table 1 shows the number of peptide and protein identifications for three replicate measurements. Generally, more peptides were identified by trypsin digestion with MS compatible detergents than either an aqueous or organic-aqueous solvent alone. In addition, trypsin digestion with MS compatible detergents yielded more peptide identifications than trypsin digestion in the presence of 2M and 8M urea ($p<0.05$). The highest number of identified peptides and the highest peptide to protein ratios were observed in 80% acetonitrile with Invitrosol (80%ACN IVS) followed by protein digestion in an aqueous buffer, Tris RPG and Tris PPS.

B. No significant non-tryptic peptides are identified in longer incubation of trypsin digestion—Incomplete digestion of proteins can limit the analysis of complex protein mixtures. To facilitate complete proteolysis of proteins long digestion times (> 5h) are often used [14]. In addition, the presence of proteolytically resistant proteins in complex protein mixtures requires vigorous denaturation and long digestion times to promote complete proteolysis. To ensure complete protein proteolysis, we allowed trypsin digestion to proceed for 16 hours. Longer incubation times with trypsin can result in alternate cleavage sites beyond lysine and arginine, consequently we examined the occurrence of cleavage at residues other then lysine and arginine as a function of the different buffer systems. We observed on average less than 10% of the total peptides were non-tryptic under all digest conditions (Table 1). The mixed organic-aqueous buffers yielded higher than average percentages (>95%) of fully tryptic peptides except 80% ACN PPS (92.9%) and 80%MeOH IVS (91.5%). In contrast to the study using purified proteins by Strader et al. [15], we found only slightly lower trypsin specificity in aqueous buffer (93.5%). A higher percentage of peptides with partial tryptic specificity was generated during digestion in aqueous buffer than the mixed the organic-aqueous system, consistent with observations by Klammer et al. [16] (Table 1).

C. MS compatible detergents increase total number of proteins identified—The total number of proteins identified in each digestion scheme is shown in Table 2. Only proteins

identified by at least two peptides are included on this list. Trypsin digestion in Tris buffer with either RPG (Tris RPG) or PPS (Tris PPS) yielded a slightly higher number of protein identifications than trypsin digestion in 80% acetonitrile with Invitrosol (p=0.07). The highest peptide to protein ratios, however, were achieved using 80% ACN IVS and 80% ACN. The peptide to protein ratio is important because a higher peptide to protein ratio reflects higher protein sequence coverage and can lead to greater confidence in protein identifications. Interestingly, trypsin digestion with MS compatible detergents in both solvent systems generated more protein identifications than Lys-C/trypsin digestion with 2M and 8M urea. For example, Tris RPG digestion generates more than a 2-fold increase in protein identification over Tris 8M urea (Table 2). This result may be explained by the known instability of urea at elevated temperature (37°C) [17] or incomplete digestion due to interference of proteolysis at a higher concentration of urea. Furthermore, there seems to be a solvent preference for the MS compatible detergents. Trypsin digestion in Tris buffer with acid labile detergents RPG and PPS resulted in more identified proteins than digestion in buffer with 80% acetonitrile (Table 2). Tris PPS digestion produced a 40.6% increase in protein identifications compared to the use of PPS in 80%ACN buffer. Aqueous Tris RPG digestion increased protein identifications by 22.7% over RPG digestion in 80%ACN. In contrast, INV digestion in aqueous Tris or 80% ACN produced the opposite result. Aqueous Tris IVS digestion produced a 30.4% decrease in protein identifications over IVS digestion in 80% ACN buffer. The impact of buffer composition in combination with MS compatible detergents has not been reported in the literature and thus these results provide valuable information to identify the optimal digest condition.

D. Comparison of uniquely identified proteins across all digestion conditions— Protein solubility varies in different buffers and in the presence of different types of detergents. Our data indicated that all three MS compatible detergents can increase protein solubility and the number of peptide and protein identifications by μLC-MS/MS. Next, we examined the similarity of protein identifications among merged triplicate experiments of each digestion condition. Systematic variations caused by data-dependent acquisition are observed among all digest conditions even after triplicate runs [18]. Nevertheless, a greater number of proteins are uniquely identified in aqueous Tris RPG digestion followed by aqueous Tris PPS digestion, and IVS digestion in 80%ACN (Figure 2). This result indicates differential protein solubilization by MS compatible detergents and suggests potential complementary applications of these detergents.

E. Peptides generated using different trypsin digestion conditions exhibit distinctly different physical properties—To analyze the characteristics of peptides generated in each proteolytic condition, we compared four parameters: isoelectric point (pI), average hydrophobicity [19], and length. While only a slight difference in pI and length was detected among peptides identified in each digest condition within the detectable m/z range (200-4000), we found interesting differences in peptide hydrophobicity. Most of the digest conditions in both the mixed organic-aqueous and aqueous systems generated relatively hydrophilic peptides (higher negative KD values) (Figure 3A). Consequently, peptides are eluted earlier in the separation gradient, and several examples of the hydrophilic peptide elution profiles are shown in Figure 3B. Interestingly, we also found several digestion conditions resulted in more hydrophobic peptides. First, IVS digestion in 80% MeOH (Ave. KD, -1.2) generated more hydrophobic peptides although fewer peptides were identified from these digestion conditions (Table 3 and Figure 3A). Several other conditions such as PPS digestion in 80%ACN (Ave. KD, -2.6), IVS digestion in 80%ACN (Ave. KD, -1.7), and aqueous Tris PPS digestion (Ave. KD, 0.5) resulted in more hydrophobic peptides and eluted later in the separation gradient (Figure 3A & C). Since the soluble fraction of cancer cell lysates was used in this study, it is less likely that only hydrophobic proteins were solubilized in conditions

which yield more hydrophobic peptides. A potential explanation for the observed differences in peptide hydrophobicity resulting from different digestion conditions with MS compatible detergents could be attributed to the structural differences among these three detergents. Each of the three detergents tested interact and denature proteins in a unique way, dictated by the structure of the detergent. Furthermore, the way in which each detergent interacts with proteins can dramatically influenced by the buffer system, resulting in different detergent-protein interactions or detergent-protein micelles. The hydrophobic portion of the protein can potentially be exposed to the solvent containing trypsin differently in each case, resulting in varying amounts of hydrophobic portion of the protein that is normally inaccessible to proteolysis. Once cleaved, each of the three detergents tested play an important role in continuing to solubilize the hydrophobic fragments and prevent the loss of these peptides through precipitation or adsorption to surfaces.

Proteolytic Digestion With MS Compatible Detergents Increase Protein Identifications By Generating Diverse Mixture Of Peptides

Traditionally, one can increase protein identifications by performing replicate analyses. However, we find that proteolytic digestion of proteins in solution with MS compatible detergents generate more diverse mixture of peptides and can contribute to a greater increase in protein identifications than technical replicates. For example, merging the data from three replicates of samples digested in trypsin with 80% acetonitrile generated 295 non-redundant protein identifications (less than 1% protein false positive rate and at least two peptide identifications per protein). However, a 1.7-fold increase in protein identifications over three replicate analyses of samples digested with trypsin in 80% acetonitrile was observed by merging the data from three replicate analyses of protein digested in aqueous Tris RPG digestion (489 non-redundant protein identifications). We demonstrated that a further increase in protein identifications (2.4-fold) over three replicate analyses of samples digested with trypsin in 80% acetonitrile can be achieved by merging the data from aqueous digestion with the data obtained from digestion with the three different MS compatible detergents (712 non-redundant protein identifications). We believe the complementary nature of MS compatible detergents creates an additional strategy to increase protein coverage in shotgun proteomics in addition to performing technical replicates. Additionally, one can customize the digestion conditions with MS compatible detergents depending on the type of characterization desired from the complex protein mixture. For example, aqueous RPG or PPS digestion can be used to achieve a higher number of protein identifications from a protein mixture. Alternatively, IVS digestion in 80% acetonitrile can be used to increase sequence coverage of the identified proteins. Lastly, aqueous PPS or PPS digestion in 80% acetonitrile can be used to increase the identification of hydrophobic peptides or to potentially increase the number of hydrophobic proteins identified.

In conclusion, the modification of trypsin digestion protocols by the addition of MS compatible detergents and organic solvents has been shown to improve proteolytic processing of complex protein mixtures. The different types of detergents have also been shown to produce variations in the numbers and types of peptides and proteins observed in LC/MS/MS experiments. These observations may have utility to increase the number of proteins identified by using combinations of detergents and organic solvents. An improvement in the identification of hydrophobic peptides was shown with the use of PPS in aqueous buffer or PPS in a buffer containing 80% acetonitrile. This observation suggests PPS may have added utility for the analysis of membrane proteins to recover more of the transmembrane fragments. Consequently, we are investigating the possibility of improving membrane protein identification by incorporating MS compatible detergents in the digestion protocol and using this class of detergents to lyse cells, extract, and solubilize proteins from cells.

Experimental Section

Materials

Invitrosol™ was purchased from Invitrogen (Carlsbad, CA). RapiGest™ SF acid labile surfactant was purchased from Waters Corp. (Milford, MA). PPS™ Slient surfactant was provided by Dr. Norris from Protein Discovery Inc. (Knoxville, TN). The proteases endoproteinase Lys-C and trypsin (modified, sequencing grade) were obtained from Roche. Other laboratory reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless noted otherwise.

Sample Preparation

Cell pellets of a mouse pancreatic cell line were provided by Dr. Cheresh at the UCSD Cancer Center. Proteins were extracted from the cell pellets using the total protein extraction kit from Biochain (Hayward, CA).

Catalase solubilization and BSA digestion with MS compatible detergents

5ug of Catalase were solubilized in 30μl of 20mM Tris-HCl buffer (pH8.0), Tris + invitrosol (IVS), Tris + PPS, Tris + RapiGest SF (RPG), and Tris + 2M urea. Samples were incubated at 37°C for 30 minutes with gently shaking to facilitate solubilization and centrifuged to remove insoluble proteins. The supernatant was heat denatured, run on a 10% acrylamide gel, and stained with Coomassie staining to visualize protein solubilization.

Trypsin digestion of BSA at a ratio of 1:50 in 20mM Tris-HCl buffer (pH8.0) with or without each MS compatible detergent was set up to visualize the digestion efficiency over 5 hours. After trypsin addition, samples were taken every hour up to 5 hours from each digestion. The digest efficiency was visualized on a 10% acrylamide gel with Coomassie staining.

Digestion protocols

15μg of cell lysates were precipitated with methanol/chloroform and resuspended in different conditions for trypsin digestion. Each digestion condition was run in triplicates.

Trypsin digestion with the mixed organic-aqueous and aqueous systems— 50μl of solvents (80% acetonitrile, 80% Methanol, 80% 2-Propanol, and 100mM Tris-HCl, pH8.0) were used to resuspend the protein pellets. All samples were sonicated for 2 hours in a water batch and proteins were digested by trypsin with gentle shaking at 37°C for 16 hours. For trypsin digestion in aqueous Tris buffer, samples were reduced in 5mM TCEP for 20 minutes at room temperature, and cysteines were alkylated in the dark in 25mM iodoacetamide (IAM) for 30 minutes at room temperature prior to the addition of trypsin. Trypsin was added to each sample at a ratio of 1:50 enzyme/protein along with 2 mM $CaCl_2$.

Trypsin digestion with the MS detergents—5ul of invitrosol (5X stock), 5ul of RapiGest SF (1% stock), and 10ul of PPS (1% stock) were used to resuspend the protein pellets. The concentration of each detergent used in this study was determined based on the maximum recommended concentration suggested by the manufactures. Then, proteins were incubated at 60°C for 5 minutes and the remaining solvent was added to the samples (50ul final volume). All samples were sonicated for 2 hours in a water batch and digested with trypsin for 16 hours at 37°C. For the aqueous digestion, samples were reduced, alkylated and digested with trypsin as described in the previous digestion condition.

Trypsin digestion with urea—Urea was added as solid to the aqueous digestion so the final concentration of urea in solution is 2M and 8M. All samples were sonicated for 2 hours

in a water batch and proteins were digested with trypsin for 16 hours at 37°C. 2M urea samples were reduced, alkylated and trypsin was added as described in the previous digestion condition. For 8M urea samples, 5ng of Lys-C were added to each sample at an enzyme/protein ratio of 1:100 and incubated for 4 hours with shaking at 37°C. Then, the 8M urea samples were diluted 1:1 with 20mM Tris, pH8.0, to a final concentration of 4M urea. Trypsin was added then an enzyme/protein ratio of 1:50 along with 2 mM $CaCl_2$.

Post-Digestion

Following digestion, all reactions were acidified with 90% formic acid (2% final) to stop the proteolysis. Samples with RapiGest SF and PPS were acidified and incubated at 37°C for additional 4 hours to facilitate the hydrolysis of detergents. Then, samples were centrifuged for 30 minutes at 14,000 rpm to remove insoluble material. The soluble peptide mixtures were collected, dried by a speedvac,a and reconstituted in 10ul of buffer A (95% H_2O, 5% acetonitrile, and 0.1% formic acid) for LC-MS/MS analysis.

Reverse phase chromatography and tandem mass spectrometry

Each peptide mixture was analyzed by automated microcaplillary liquid chromatography-tandem mass spectrometry. Fused silica capillaries (100μm i.d.) were pulled using a P-2000 CO2 laser puller (Sutter Instruments, Novato, CA) to a 5μm i.d. tip and packed with 12 cm of 3μm Aqua C18 material (Phenomenex, Ventura, CA) using a pressure bomb. The column was equilibrated in buffer A, and the peptide mixture was loaded onto the column using a pressure bomb. This column was then placed in-line with a Quaternary Agilent (Foster City, CA) 1100 series HPLC pump. The HPLC pump flows at 0.2 μl/min, and the flow rate to the electrospray tip is reduced to ~ 200-300 nl/min. The HPLC separation was provided by a gradient between Buffer A (95% water, 5% acetonitrile, and 0.1% formic acid) and Buffer B (20% water, 80% acetonitrile, and 0.1% formic acid). The HPLC gradient was held constant at 100% buffer A for 5 minutes and then changed from 0% buffer B to 10% buffer B over 5 minutes followed by a 90-min gradient from 10% buffer B to 55% buffer B. Then, the gradient was switched from 55% to 100% buffer B over 5 minutes and held constant for 5 minutes. Finally, the gradient was changed from 100% buffer B to 100% buffer A over 5 minutes, and then held constant at 100% buffer A for 5 more minutes. The application of a 2.5 kV distal voltage electrosprayed the eluted peptides directly into an LTQ ion trap mass spectrometer equipped with a nano-LC electrospray ionization source (ThermoFinningan, San Jose, CA). Full masses (MS/MS) spectra were recorded on the peptides over a 400-2000 m/z range, followed by three tandem mass (MS/MS) events sequentially generated in a data-dependent manner on the first, second, and third most intense ions selected from the full MS spectrum (at 35% collision energy). Mass spectrometer scan functions, and HPLC solvent gradients were controlled by the Xcalibur data system (ThermoFinnigan, San Jose, CA).

Database search and interpretation of MS/MS datasets

Tandem mass spectra were extracted from raw files, and a binary classifier [20] - previously trained on a manually validated data set - was used to remove the low quality MS/MS spectra. The remaining spectra were searched against a *Mus Musculus* protein database containing 50,370 protein sequences downloaded as FASTA-formatted sequences from EBI-IPI (database version 3.11, released on October 11, 2005), and 124 common contaminant proteins, for a total of 50,494 target database sequences. To calculate confidence levels and false positive rates, we used a decoy database containing the reverse sequences of the 50,494 proteins appended to the target database [21], and the SEQUEST algorithm [22] to find the best matching sequences from the combined database. SEQUEST searches were done on an Intel Xeon 80-processor cluster running under the Linux operating system. The peptide mass search tolerance was set to 3 Da. No differential modifications were considered. For the aqueous digestion, the mass of

the amino acid Cysteine was statically modified by +57 Da, due to carboxyamidomethylation of the sample. No enzymatic cleavage conditions were imposed on the database search, so the search space included all candidate peptides whose theoretical mass fell within the 3 Da mass tolerance window, despite their tryptic status.

The validity of peptide/spectrum matches was assessed in DTASelect2[4] using SEQUEST-defined parameters, the cross-correlation score (XCorr) and normalized difference in cross-correlation scores (DeltaCN). The search results were grouped by charge state (+1, +2, and +3) and tryptic status (fully tryptic, half-tryptic, and non-tryptic), resulting in 9 distinct sub-groups. In each one of the sub-groups, the distribution of XCorr and DeltaCN values for (a) direct and (b) decoy database hits was obtained, and the two subsets were separated by quadratic discriminant analysis. Outlier points in the two distributions (for example, matches with very low Xcorr but very high DeltaCN) were discarded. Full separation of the direct and decoy subsets is not generally possible; therefore, the discriminant score was set such that a false positive rate of 5% was determined based on the number of accepted decoy database peptides. This procedure was independently performed on each data subset, resulting in a false positive rate independent of tryptic status or charge state.

In addition, a minimum sequence length of 7 amino acid residues was required, and each protein on the list was supported by at least two peptide identifications unless specified. These additional requirements – especially the latter - resulted in the elimination of most decoy database and false positive hits, as these tended to be overwhelmingly present as proteins identified by single peptide matches. After this last filtering step, the false identification rate was reduced to below 1%.

Finally, the Kyte-Doolittle hydropathy score and the isoelectric point (pI) of all the peptides passed the filters were also calculated and reported in the DTASelect2 output.

Acknowledgements

This research was supported by an NIAID sub-contract grant UCSD/MCB0237059 and partially supported by NIH P41RR11823 and NIH 5R01 MH067880.

References

1. Du Y, Parks BA, Sohn S, Kwast KE, Kelleher NL. Top-down approaches for measuring expression ratios of intact yeast proteins using fourier transform mass spectrometry. Anal Chem 2006;78(3):686–94. [PubMed: 16448040]
2. Meng F, 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(13):2923–9. [PubMed: 12141648]
3. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd. Direct analysis of protein complexes using mass spectrometry. Nature Biotechnology 1999;17(7):676–82.
4. Tabb DL, McDonald WH, Yates JR r. DTASelect and Contrast: Tools for Assembling and Comparing Protein Identification from Shotgun Proteomics. J Proteome Res 2002;1:21–26. [PubMed: 12643522]
5. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 1999;20(18):3551–67. [PubMed: 10612281]
6. Lenard J, Singer SJ. Protein Conformation in Cell Membrane Preparations as Studied by Optical Rotatory Dispersion and Circular Dichroism. Proc Natl Acad Sci U S A 1966;56(6):1828–1835. [PubMed: 16591427]
7. Gordon JA, Jencks WP. The relationship of structure to the effectiveness of denaturing agents for proteins. Biochemistry 1963;2:47–57. [PubMed: 13949214]

8. Rajagopalan KV, Fridovich I, Handler P. Competitive inhibition of enzyme activity by urea. J Biol Chem 1961;236:1059–65. [PubMed: 13739212]

9. J, CB.; J, LJ.; K, BE.; G, HR.; J, D.; J, L.; P, BES. A novel SDS analog compatible with MS analysis of proteins and peptides. Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics; Dallas, TX. 13-17 June; 1999.

10. Nomura E, Katsuta K, Ueda T, Toriyama M, Mori T, Inagaki N. Acid-labile surfactant improves in-sodium dodecyl sulfate polyacrylamide gel protein digestion for matrix-assisted laser desorption/ionization mass spectrometric peptide mapping. J Mass Spectrom 2004;39(2):202–7. [PubMed: 14991690]

11. Yu YQ, Gilar M, Lee PJ, Bouvier ES, Gebler JC. Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. Anal Chem 2003;75(21):6023–8. [PubMed: 14588046]

12. Zybailov B, Coleman MK, Florens L, Washburn MP. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. Anal Chem 2005;77(19):6218–24. [PubMed: 16194081]

13. Russell WK, Park ZY, Russell DH. Proteolysis in mixed organic-aqueous solvent systems: applications for peptide mass mapping using mass spectrometry. Anal Chem 2001;73(11):2682–5. [PubMed: 11403317]

14. Lundell N, Schreitmuller T. Sample preparation for peptide mapping--A pharmaceutical quality-control perspective. Anal Biochem 1999;266(1):31–47. [PubMed: 9887211]

15. Strader MB, Tabb DL, Hervey WJ, Pan C, Hurst GB. Efficient and specific trypsin digestion of microgram to nanogram quantities of proteins in organic-aqueous solvent systems. Anal Chem 2006;78(1):125–34. [PubMed: 16383319]

16. Klammer AA, MacCoss MJ. Effects of modified digestion schemes on the identification of proteins from complex mixtures. J Proteome Res 2006;5(3):695–700. [PubMed: 16512685]

17. Lippincott J, Apostol I. Carbamylation of cysteine: a potential artifact in peptide mapping of hemoglobins in the presence of urea. Anal Biochem 1999;267(1):57–64. [PubMed: 9918655]

18. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem 2004;76(14):4193–201. [PubMed: 15253663]

19. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 1982;157:105–132. [PubMed: 7108955]

20. Bern M, Goldberg D, McDonald WH, Yates JR 3rd. Automatic quality assessment of Peptide tandem mass spectra. Bioinformatics 2004;20:I49–I54. [PubMed: 15262780]

21. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res 2003;2(1):43–50. [PubMed: 12643542]

22. Yates JR III. Database searching using mass spectrometry data. Electrophoresis 1998;19:893–900. [PubMed: 9638935]

Figure 1A.

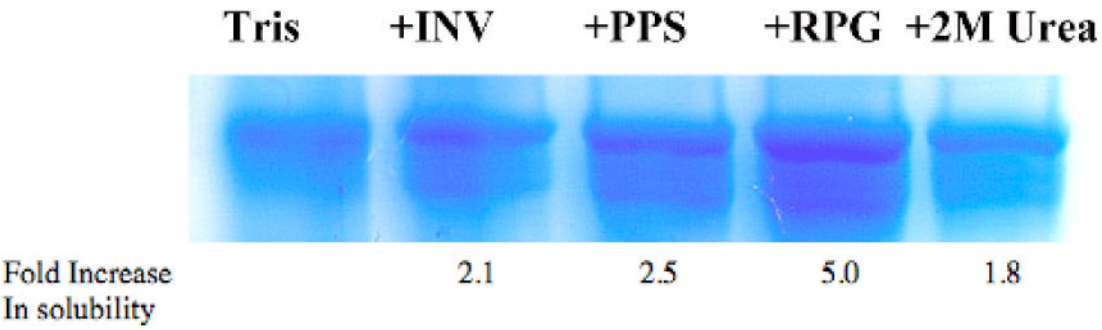


Figure 1B.

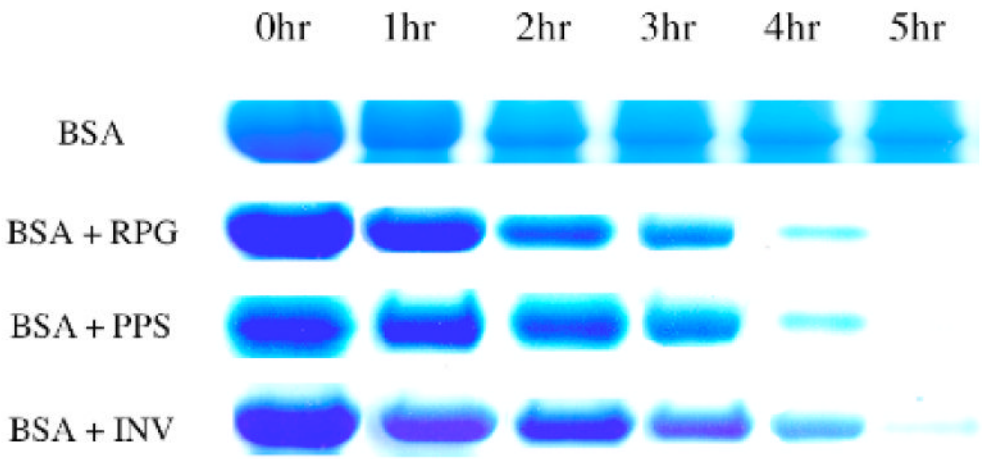
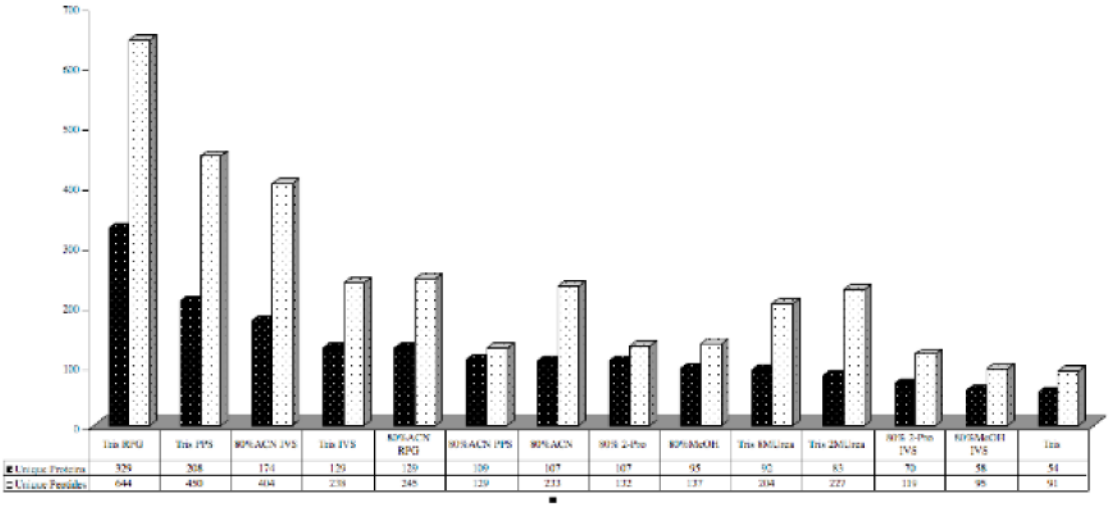


Figure 1. MS compatible detergents improve protein solubility and digest efficiency in solution

5μg of Catalase were resuspended in 20mM Tris-HCl, pH8.0 buffer with or without invitrosol (IVS), PPS, RapiGest SF (RPG), and 2M urea (panel A). More protein was solubilized in MS compatible detergents and 2M urea than the buffer alone. In panel B, BSA was digested with or without each detergent in 20mM Tris-HCl, pH8.0 buffer over 5 hours at 37°C. While a substantial amount of the full-length BSA remained in solution from the trypsin digestion in Tris buffer alone, increased proteolysis of BSA in the presence of each MS compatible detergent was observed based on disappearance of the full-length BSA over time visualized by coomassie stained 10% acrylamide gels.



ACN = acetonitrile, IVS = Invitrosol (invitrogen), RPG = Rapigest (Waters), PPS (Protein Discovery), MeOH = Methanol, 2-Pro = 2-propanol, Tris = 100mM Tris-HCl, pH 7.5

* Unique protein and peptide identifications in Tris RPG are significantly higher ($p<0.05$) than all other digestions.

Figure 2. Comparison of uniquely identified proteins and peptides across all digestion conditions

Protein and peptide identifications from triplicate experiments were merged and filtered based on a minimum two peptides and less than 5% protein false positive rate. Then, proteins identified from each digestion condition were compared with each other and proteins identified exclusively in each digest condition are shown in figure 2. The results are ranked by uniquely identified proteins from the most (left) to the least (right).

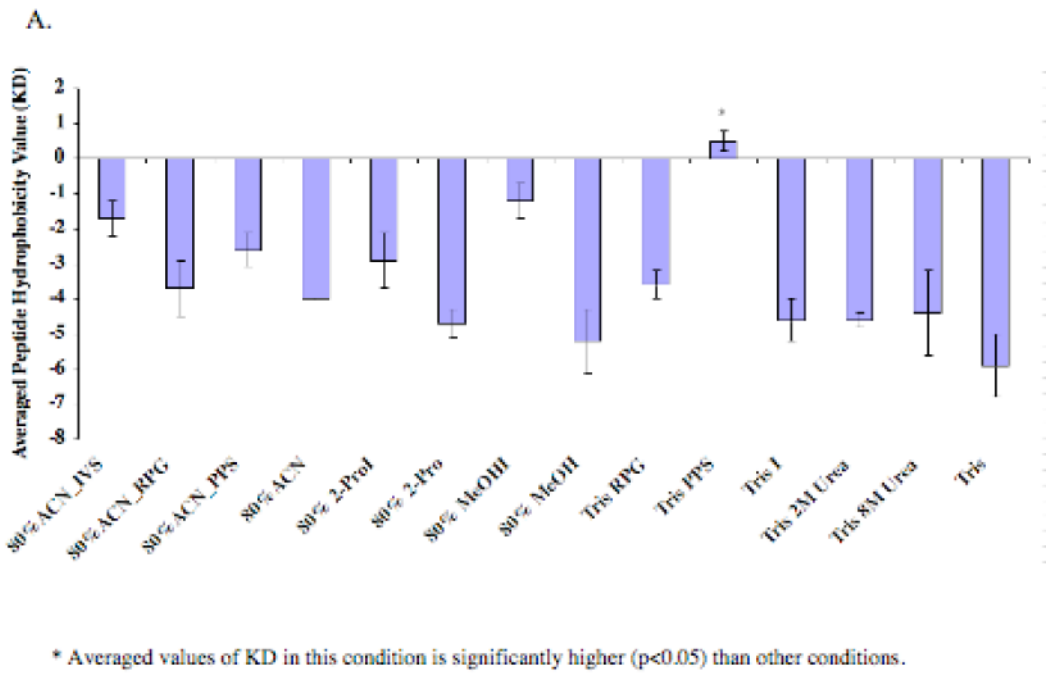


Figure 3. Comparison of the averaged hydrophobicity of peptides identified in each digestion condition

The predicted KD of peptides identified in each digestion condition is shown in figure 3A. The peptide is predicted to be more hydrophilic toward the negative value and more hydrophobic toward the positive value. The elution profiles of peptides identified over the elution gradient (represented by scan number) for six digestion conditions are shown. The elution profiles of more hydrophilic peptides, including aqueous Tris IVS digestion (Ave. KD, -4.6), RPG digestion in 80%ACN (Ave. KD, -3.6), and aqueous Tris digestion in 8MUrea (Ave. KD, -4.4), are shown in panel B. The elution profiles of more hydrophobic peptides, including PPS digestion in 80%ACN (Ave. KD, -2.6), IVS digestion in 80%ACN (Ave. KD, -1.7), and aqueous Tris PPS digestion (Ave. KD, 0.5), are shown in panel C.

Table 1

Comparison of trypsin digest with MS compatible detergents in mixed organic-aqueous and aqueous systems

Cytoplasmic fraction of the mammalian cell lysate was digested with or without MS compatible detergents in either mixed organic-aqueous or aqueous systems to determine the effect of various conditions on trypsin proteolysis. Each condition was repeated three times and the number of identified peptides and proteins are shown as Rep 1, 2, and 3. Peptide per protein ratio is determined using the average value from the triplicates, and the average value of tryptic status is shown in the table. Protein identification was based on minimum one peptide for the purpose of peptide analyses and less than 5% protein false positive rate.

	# of peptides identified				# of protiens identified							
	Rep1	Rep2	Rep3	Average	Rep1	Rep2	Rep3	Average	Peptide/Protein	% Fully Typtic	% Half Typtic	% Non Tryptic
Organic												
80%ACN IVS	1731	1516	1504	**1584**	431	350	384	**388**	4.1	96.0	4.0	0.0
80%ACN RPG	1142	1214	1014	**1123**	398	391	361	**383**	2.9	97.2	2.6	0.2
80%ACN PPS	1185	1020	1100	**1102**	350	316	300	**322**	3.4	92.9	0.9	6.2
80%ACN	948	1064	888	**967**	313	349	304	**322**	3.0	98.6	1.1	0.3
80% 2-Pro IVS	763	842	812	**806**	246	265	260	**257**	3.1	98.5	1.1	0.4
80% 2-Pro	692	773	773	**746**	251	278	230	**253**	2.9	98.1	1.5	0.3
8% MeOH IVS	283	280	243	**269**	192	164	203	**186**	1.4	91.5	6.7	1.8
80% MeOH	200	171	171	**181**	140	160	171	**157**	1.2	95.6	1.0	3.4
Aqueous												
Tris RPG	1207	1246	1368	**1274**	425	422	472	**440**	2.9	90.6	8.1	1.3
Tris PPS	1093	1319	1159	**1190**	476	469	400	**448**	2.7	96.5	3.4	0.2
Tris IVS	1010	1084	1095	**1063**	326	322	323	**324**	3.3	91.7	7.7	0.6
Tris 2M Urea	928	989	987	**968**	328	341	323	**331**	2.9	92.7	7.0	0.3
Tris 8M Urea	590	621	606	**606**	212	220	208	**213**	2.8	93.1	6.5	0.4
Tris	225	234	252	**237**	131	109	121	**120**	2.0	96.3	2.1	1.6

ACN = acetonitrile, IVS = Invitrosol (invitrogen), RPG = Rapigest (Waters), PPS (Protein Discovery), MeOH = Methanol, 2-Pro = 2-propanol, Tris = 100mM Tris-HCl, pH 7.5

* Peptide identification in 80% ACN IVS is significantly higher ($p<0.05$) than other organic digestions. Peptide identification in Tris RPG is significantly higher ($p<0.05$) than other aqueous digestions.

Table 2

Total protein and peptide identifications of trypsin digestion with MS compatible detergents in mixed organic-aqueous and aqueous systems

Protein and peptide identifications from triplicate experiments are merged and filtered based on minimum two peptide for higher confidence and less than 5% protein false positive rate. The results are ranked from the most protein identified to least protein identified. Peptide to protein ratio is generated by dividing the total number of peptides identified over the total number of proteins identified.

	Proteins	Peptides	Peptides/Protein
Tris RPG	453	2358	5.2
Tris PPS	429	2164	5.0
80% ACN IVS	395	2437	6.2
80% ACN RPG	350	1998	5.7
TRIS 2M Urea	306	1741	5.7
80% ACN	294	1751	6.0
Tris IVS	275	1577	5.7
80% ACN PPS	255	1271	5.0
80% 2-Pro	206	1070	5.2
Tris 8M Urea	199	1148	5.8
80% 2-Pro IVS	181	1071	5.9
Tris	149	773	5.2
80% MeOH IVS	128	531	4.1
80% MeOH	115	464	4.0

Table 3

Characteristics of peptides generated by trypsin digestion with MS compatible detergents in mixed organic-aqueous and aqueous systems

Characteristics of peptides generated by trypsin digestion with or without MS compatible detergents in either mixed organic-aqueous or aqueous systems were determined based on the predicted isoelectric point (pI), averaged hydrophobicity (KD), and length. Averaged values of each parameter and standard deviations (STDEV) were calculated from triplicate experiments.

	Ave. pI (STDEV)	Ave. KD (STDEV)	Ave. Length (STDEV)
80%ACN_IVS	5.4 (0.2)	-1.7 (0.5)	18.7 (1.0)
80%ACN_RPG	5.4 (0.1)	-3.7 (0.8)	15.9 (1.1)
80%ACN_PPS	5.7 (0.1)	-2.6 (0.5)	15.5 (0.6)
80%ACN	5.7 (0.2)	-4.0 (0.0)	14.7 (0.9)
80% 2-ProI	5.7 (0.3)	-2.9 (0.8)	16.3 (0.4)
80% 2-Pro	5.4 (0.1)	-4.7 (0.4)	14.1 (0.0)
80% MeOHI	5.7 (0.1)	-1.2 (0.5)	18.1 (1.4)
80% MeOH	5.7 (0.5)	-5.2 (0.9)	15.6 (0.6)
Tris RPG	5.2 (0.0)	-3.6 (0.4)	16.2 (1.4)
Tris PPS	5.5 (0.2)	0.5 (1.5)	16.1 (1.6)
Tris I	5.6 (0.6)	-4.6 (0.6)	18.6 (1.8)
Tris 2M Urea	5.3 (0.1)	-4.6 (0.2)	18.8 (0.6)
Tris 8M Urea	5.7 (0.2)	-4.4 (1.2)	17.2 (2.4)
Tris	5.3 (0.1)	-5.9 (0.9)	16.8 (0.7)